Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

Studies of Pupillary Conditioning: F. A. YOUNG 97
"Backward" Associations in Transfer and Learning: B. B. MURDOCK, JR. 111
Probability Learning Under Equivalent Data Collection Methods: S. S. KOMORITA 115
Expectation and Resistance to Extinction of a Lever-Pulling Response as a Function of Percentage of Reinforcement and Number of Acquisition Trials: D. J. LEWIS AND C. P. DUNCAN 121
Effects of Mediated Associations on Transfer in Paired-Associate Learning: K. J. NORCROSS AND C. C. SPIKER 129
Transfer Effects From a Single to a Double Integral Tracking System: G. E. BRIGGS, P. M. FITTS, AND H. P. BAHRICK 135
Some Informational Aspects of Form Discrimination: G. K. KRULEE 143
Assimilation and Contrast Effects of Anchoring Stimuli on Judgments: M. SHERIF, D. TAUB, AND C. I. HOVLAND 150
Achievement Drive and Habitual Modes of Task Approach as Factors in Skill Transfer: G. H. MILES 156
Influence of Task Complexity and Instructions Upon Simple and Discrimination Reaction Times: J. B. SIDOWSKI, R. MORGAN, AND G. ECKSTRAND 163
Effects of Task Difficulty, Stimulus Similarity, and Type of Response on Stimulus Predifferentiation: B. B. MURDOCK, JR. 167
Motor Conflict Behavior as a Function of Motivation and Amount of Training: B. G. ANDREAS 173
Perceptual Efficiency as Related to Induced Muscular Effort and Manifest Anxiety: M. F. SHORE 179
Evidence That the Secondary Reinforcing Stimulus Must be Discriminated: F. J. MCGUIGAN AND F. CROCKETT 184
Muscle Responses and Their Relation to Rote Learning: R. N. BERRY AND R. C. DAVIS 188
Cue and Contextual Stimulus Intensity in Discrimination Learning: J. E. NYGAARD 195
Shifts in Connotative Meaning of Words as a Function of Previous Restrictive Experience: M. S. MAYZNER AND M. E. TRESSELT 200

American Psychological Association

Vol. 55 No. 2 February 1958

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Circulation Manager
SADIE J. DOYLE, Editorial Assistant

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor: Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices in this Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1958 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1958 by the American Psychological Association, Inc.

This issue of the *Journal of Experimental Psychology* is one of a series of issues with an increased number of pages so as to reduce the present publication backlog. This increase in pages has been made possible by a grant from the National Science Foundation to the American Psychological Association. The grant was given for the specific purpose of reducing publication lag; it carries the provision that editorial standards of the journal will remain unchanged.

Journal of
Experimental Psychology

VOL. 55, NO. 2 FEBRUARY, 1958

STUDIES OF PUPILLARY CONDITIONING[1]

FRANCIS A. YOUNG

State College of Washington

Baker (**2**) and Hudgins (**5**) have reported conditioning of pupillary reflexes, but a number of others (**3, 4, 8, 9, 10, 11, 14, 15**) have been unable to verify these findings. The four studies reported here include exact repetitions of Baker's and of Hudgins' experiments and attempts to condition the pupil using stimuli other than light as the UCS. In no case was pupillary conditioning induced.

EXPERIMENT I

Apparatus and method.—Except for the substitution of an infrared motion picture camera (**12**) for the telescope and kymograph, Hudgins' (**5**) early study of pupillary conditioning was repeated exactly. A duplicate of this stimulus apparatus was built according to his description and drawings (**5**, pp. 3–5).

A 100-w. light bulb was mounted in the base of a box enclosure and in front of a silvered concave reflector at a distance of 136 cm. from the eye of *S*. The light passed through a lens which focused it on a 90° prism. From here, the light passed into the eye of *S*. The intensity of light was controlled by an iris diaphragm, which was placed immediately beneath the lens. Two intensities of light were used as the UCS. The training light was the maximum intensity emitted by the bulb through an orifice 28 mm. in diameter. The test light came from the same source, but was emitted through an orifice 4 mm. in diameter. The photometric measurement of these two lights was: training light, 37.7 meter-candles; test light, 3.5 meter-candles. The training light was used throughout all the conditioning procedure, and the test light in all tests.

The fixation object, consisting of two fine crossed wires, was placed immediately above the light bulb. As one looked into the apparatus from *S*'s seat, the light appeared as a brilliant round disc on which was imposed the fixation cross, dividing the disc into quadrants. As the orifice of the diaphragm was made larger or smaller, the intersection of the two wires remained in the center of the disc.

The bell was a 7 cm. doorbell with a damping arrangement to prevent the continuation of the tone after the circuit was broken. Hudgins' doorbell was operated by a 2-v. storage cell; whereas, the bell in the repeat study was operated on 8 v. AC. Hudgins presented no calibration of the loudness or timbre of his bell.[2] The bell was placed in the top of the box, immediately in front of *S* and slightly above the horizontal plane of his eyes. Although Hudgins did not mention a headrest, one was used in this study.

The *S*s in the repeat experiment were seven men and six women, all between the ages of 20 and 35. Hudgins (**5**) used 14 *S*s of unspecified age and sex.

[1] This investigation was supported in part by a grant-in-aid from the National Science Foundation, and in part by funds provided for biological and medical research by the State of Washington, Initiative Measure No. 171.

[2] A personal communication from Dr. Hunter indicated that several different types of bells and buzzers were tried and the one finally used was quite loud and raucous. The bell used in the repeat study was both. He also wrote that Hudgins had used a headrest with the apparatus.

The conditioning procedure was the same as that used by Hudgins. The training-light and bell circuits were closed simultaneously by *E* (a timer in this study). The stimuli remained on for 3–4 sec. (4 sec. in this study). This was followed by 10 sec. (11 sec. here) of darkness and silence. Then, the trial was repeated. Rest periods of 3 min. were allowed at the end of every 30 trials. The training was divided into two periods. The first day, 125 trials were given, and tests were conducted at the end of this period. On the following day, the training continued until 80 to 100 additional trials (90 in this study) had been given, and a final test for this step was made at the end of this series. The method of testing for the conditioned effect of the bell was in every respect identical with the method used in the initial measurements. The initial effect of the bell was determined as follows: the test light was turned on and the pupil was measured as it stabilized. Then, *E* rang the bell for 5 sec. and again measured the pupil.

In this study, pictures were taken at the rate of 4 per sec. for 20 sec., the camera being started .5 sec. before the bell. All data represent the horizontal pupil diameters of *S*s at .25-sec. intervals.

Results.—As noted above, Hudgins did not indicate the sex distribution of his *S*s although at least two, the secretaries, were presumably female, nor did he distinguish between men and women in reporting his results. In the experiments reported here, there were consistent sex differences, and for this reason, the results of repeating Hudgins' experiment in men and women are presented separately in Fig. 1 and 2, respectively.

The pretest curve in Fig. 1 indicates that the pupil diameter in male *S*s increased rapidly; then remained constant for a period about twice as long as the duration of the bell, and then gradually decreased to a leveling-off point considerably above the starting point. The plateau during and after the ringing of the bell is probably exaggerated since the damping device used in this study did not instantane-

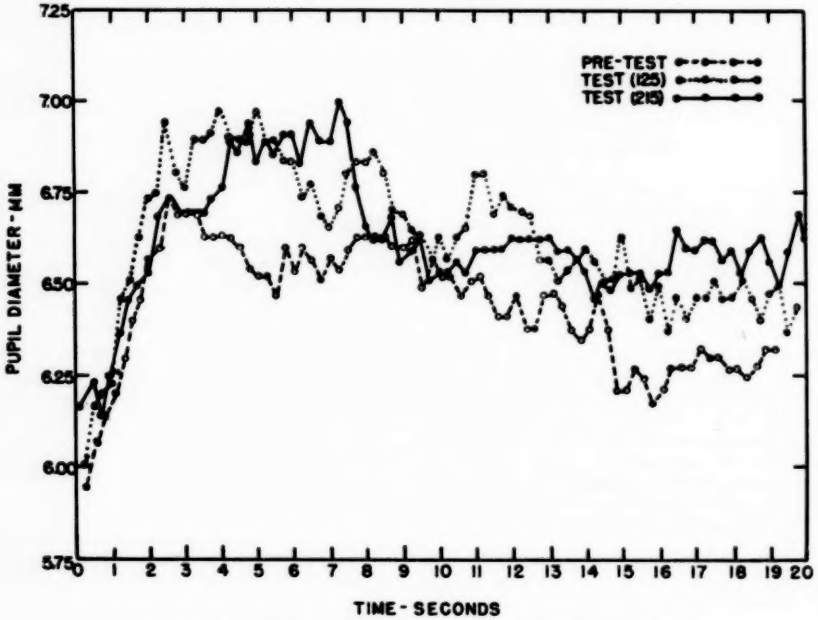

FIG. 1. The pupillary response of male *S*s to a bell (CS) before and after 125 and 215 pairings with a light (UCS).

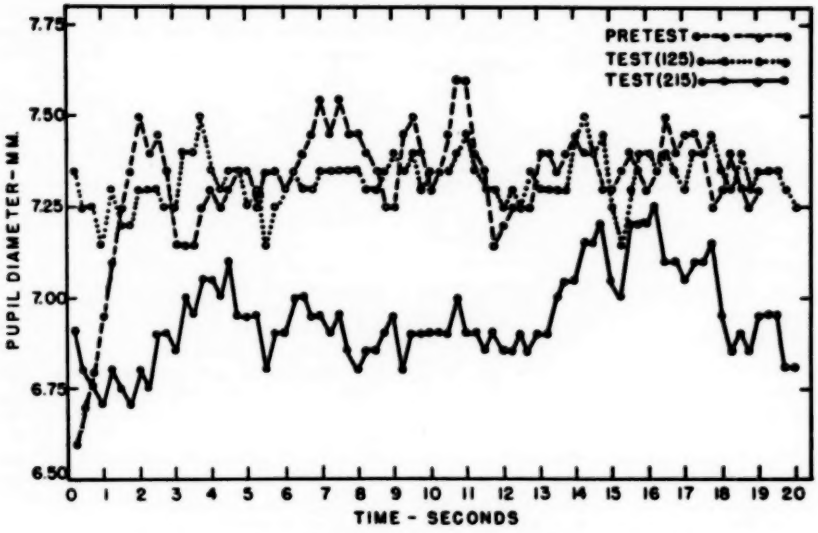

FIG. 2. The pupillary response of females to a bell (CS) before and after 125 and 215 pairings with a light (UCS).

ously stop all bell sounds. The general response to the bell was similar to that Hudgins found in one *S* (**5,** Fig. 2., and also, Fig. 7 of this paper), namely, a dilation of about .75 mm. The time relationships and the shapes of the curves are quite different, with Hudgins' *S* showing a slower increase and a more rapid decrease in pupil diameter. However, in both studies the initial UCR to the bell stimulus was a pupil dilation.

The tests for conditioning in the male *Ss* are presented in Fig. 1 as Test 125, following 125 conditioning trials, and Test 215, following 215 conditioning trials. These conditioning trials attempted to bring about a pupillary constriction in response to the ringing of the bell since the UCS was an increase in the intensity of the light. It is evident that these two conditioning-test curves, showing dilation, are almost identical, and further, that they are very similar to the curve obtained as the original UCR to the bell alone. These curves are entirely dissimilar from the curves presented by Hudgins (**5,** Fig. 3) as the CR to the bell since his curves show a contraction, .75 mm. The results presented here do not indicate a conditioned constriction.

As with the curves obtained on male *Ss*, the curves for female *Ss* (Fig. 2) show no evidence of conditioned constriction. There are several differences, however, between these curves and those of the men. The pretest curve, or initial response to the bell alone, shows the same rapid increase in diameter, but it does not show the ensuing slow decrease as strikingly as does the male curve. Further, the conditioning-test curves do not rise rapidly but seem to fluctuate about a constant straight line. The over-all smaller pupil diameters shown in the curve

for the final conditioning test and the sex differences will be discussed in more detail following the presentation of the results of Exp. IV.

Experiment II

Some difficulty was encountered in the attempt to duplicate Baker's (**2**) experiment, and personal correspondence was resorted to in order to insure that the duplication be as accurate as possible.[3] The following description of apparatus and procedure is taken from the original paper and the later correspondence, with any differences in procedure clearly indicated.

Apparatus and method.—The auditory stimulus was administered to *S* by means of a pair of ordinary headphones connected to an audiometer and equipped with a pair of rubber earpads to reduce extraneous noises. The audiometer in both studies was of the inductorium type used by Seashore (**7**), and was calibrated on two observers to insure that the distance between taps covered slightly less than a j.n.d. for these *O*s. The tap steps below the thresholds of these *O*s were the same as those in the Seashore Audiometer, and the phrase "two audiometer units below the subject's conscious threshold" means that after *S*'s limen had been determined, the slide was moved two tap steps below the limen tap. The input of the audiometer was fed through an interrupter.

Baker used a pupillometer which recorded continuously on a kymograph. Readings were taken from this continuous record at intervals of 1 sec. and averaged for all *S*s. In the repeat study, the infrared camera was substituted for the pupillometer and kymograph.

The UCS for changes in pupillary diameter was variation in the intensity of light furnished by a 60-w. electric lamp. Illumination of *S*'s eye could be varied between 3.5 and 20 ft.-candles by means of a rheostat (in this study, a Variac) introduced into the current supply of the lamp.

In Baker's study, *E* and *S* were separated from each other by a 4 ft. × 5 ft. black screen through which extended the objective end of the pupillometer. In the repeat study *S* sat in a small (4 ft. × 4 ft. × 7 ft.), light-tight booth which was painted flat black inside. The camera was mounted outside with the lens projecting through the front of the booth. The *S* sat in a straight-backed chair (or stool, here) with his head held firmly in place by a headrest and an adjustable chin support. A 5½ in. × 4½ in. milk-glass window in the screen (front of the booth) was situated at about the level of *S*'s chin, and admitted the light which served as the UCS.

When *S*'s head was in place in the apparatus, he could see a reflection of his own eye in the first lens of the pupillometer (camera). Since this mirroring surface was convex, the reflection was small and made a satisfactory fixation object. It was about 7½ in. from *S*'s eye, and was in such a position that he had to keep his eye entirely open without allowing the lid to droop and hinder *E*'s vision.

In the training of pupillary constriction, the UCS was an increase in light intensity from a base of 3.5 ft.-candles to a maximum intensity of 20.0 ft.-candles. In the training of dilation, the UCS was a decrease in light intensity from 20.0 to 3.5 ft.-candles. In the testing procedures it was found advisable to use a light intensity somewhere between these extremes, and 8.2 ft.-candles was arbitrarily chosen as an intensity which allowed change of the pupil in either direction. All tests were run with the light at this middle intensity.

During each *S*'s first experience in the experimental situation, the conscious threshold-for-the-stimulus was carefully determined by the same method Baker had used in another study (**1**). This method consisted of presenting stimuli over a wide range of intensities, and so discovering the intensity-level below which *S* could not perceive consciously the occurrence of the stimulus. In subsequent sessions, only brief examinations were made to determine whether the limen had shifted. The intensity of the stimulus was increased and decreased gradually, and *S* was asked to report the point at which he was just able to hear the sound. The *S*'s performance was then compared with that in his initial test, and adjustment was made when indicated.

In the determination of the limen, *S*'s surroundings duplicated exactly the conditions which would later appear in the conditioning procedure. His posture was the same, the sounds made by the Harvard clock and the time markers were the same, and the hum of the motor used to drive the audiometer brush was the same. In the repeat study, the audiometer brush motor was in the booth, but the rest of the apparatus was outside the booth, which was partially soundproof.

During the conditioning experiments frequent

[3] The correspondence with Dr. Baker was designed to clarify the nature and use of the audiometer, especially the audiometer units.

TABLE 1

ORDER OF TESTS IN EXP. II

Group A	Group B
1. Test-O (control test-before-training)	1. Test-sub-2 (control test-before-training)
2. Test-sub-2 (control test-before-training)	2. Test-O (control test-before-training)
3. Test-O (control test-before-training)	3. Test-sub-2 (control test-before-training)
4. Train-sub-2 (conditioning)	4. Train-O (conditioning control)
5. Test-O (control test)	5. Test-sub-2 (control test)
6. Test-sub-2 (test-for-conditioning)	6. Test-O (conditioning control)
7. Train-O (conditioning control)	7. Train-sub-2 (conditioning)
8. Test-O (conditioning control)	8. Test-sub-2 (conditioning test)
9. Test-sub-2 (conditioning test)	9. Test-O (conditioning control)
10. Test-O (conditioning control)	10. Test-sub-2 (conditioning test)

rests were given to *S* in order to relieve fatigue. Usually, six consecutive training combinations were completed between rest periods.

The CS was presented for 19 sec., and the UCS overlapped the subliminal sound for the last 10 sec. of the 19-sec. period. Thus the procedure for a single training combination was as follows: (*a*) CS (sound) presented, continued alone for 9 sec. (*b*) UCS (light) presented 9 sec. after start of CS; both stimuli continued together for 10 sec. (*c*) Both stimuli stopped 19 sec. after beginning of the presentation.

In the testing procedures, the CS was presented for 15 sec. without reinforcement by the UCS. In both the training and the testing procedure, the intervals between applications were irregularly varied between 10 and 30 sec. according to a chance distribution.

After determination of the limen, one of the procedures shown in Table 1 was used for each *S*. It will be noted the two procedures reverse the order of periods in order to control any possible effect of that factor on the results.

The testing and training with the CS at zero intensity (Test-O and Train-O in Groups A and B) were for control purposes. Since the intensity of the CS was so low (Test-sub-2 and Train-sub-2 with the stimulus two tap steps below limen) as to be unperceived by *S*, extreme care had to be taken to control all incidental stimuli. If conditioning appears with the CS at the intensity sub-2, but does not appear with the CS at zero intensity, then such conditioning cannot be attributed to any other CS than the one designated.

The above description, which is taken directly from Sections II, III, and IV of Baker's monograph (**2**), does not indicate the number of training combinations which were presented during the Train-O and Train-sub-2 sessions, although Section IV deals with the conditioning of the response. Section V, while not indicating specifically the number of combinations presented, does contain results obtained for 20 combinations on 16 *S*s, 8 of whom were trained in the direction of dilation and 8 in the direction of constriction.

Since both Sections IV and V were repeated, 20 training combinations were presented under the constriction and the dilation conditions at both Train-O and Train-sub-2. Baker used four *S*s for constriction and three for dilation when conditioning the response, and eight in each group for studying the course of conditioning. No mention was made of sex or the relationship between the *S*s used in the two experiments. In the repeat study, five males were used for constriction training and five different males for dilation training. The two sections of Baker's study dealing with conditioning and the course of conditioning were repeated in the same experimental sessions.

Results.—The results of the repetition of both sections are presented in Fig 3. Baker's Fig. 1 (**2**) is included in Fig. 3 for comparison. The control-test curve for the repeat study represents the average of the response to the Test-sub-2 (control test-before-training) of Group A and to the two parallel control tests of Group B. Baker did not indicate the source of his control data. The test curves for constriction and dilation in the repeat study represent the averaged responses on the test trial following 20 training combinations. Baker did not mention the number of training combinations preceding his test curves.

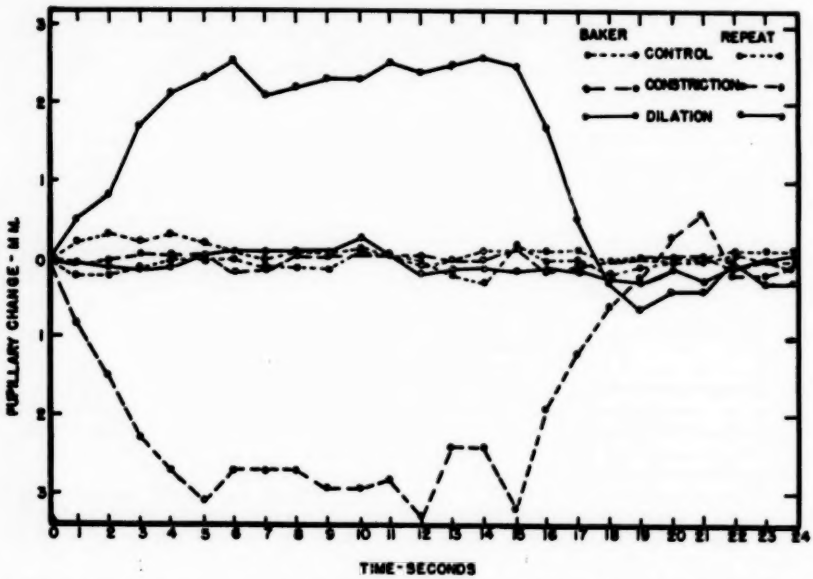

FIG. 3. The pupillary response of male *Ss* before and after 20 constriction and dilation training trials with a subliminal auditory stimulus together with the comparable curves from Baker's monograph.

The height of the curve at the beginning of the period was arbitrarily taken as zero. The height of the line above or below this point was measured at 1-sec. intervals throughout the period. These measures were then averaged for all *Ss* for each progressive second to obtain the figures upon which the curves are based. The results of the controls, in which the CS intensity was zero, are not included in the figure. Without any exception, the test curves showed absolutely no evidence of conditioning.

Discussion.—It is clearly evident from Fig. 3 that Baker's control curve and the control curve obtained in the repeat study do not differ. Nor is there any difference between these control curves and the dilation and constriction curves obtained in the repeat study. Baker's curves which indicate conditioned dilation and constriction were not verified in the repeat study. Possible explanations of this discrepancy are discussed below.

EXPERIMENT III

All known previous studies of pupillary conditioning have involved variation in light intensity as the UCS. Since light is only one of many stimuli (**6**) which will bring about a change in pupil diameter, it seemed feasible to utilize another effective stimulus as the UCS in attempting pupillary conditioning. While the pupil will respond to drugs, fatigue, accommodation and convergence, emotional states, loud sounds, etc., not all of these responses have the characteristics required for UCS. Three of these, convergence and accommodation, loud sounds, and emotional state as represented by the

response to a mild electric shock, are easily controlled, and elicit a relatively small but consistent pupillary response. Convergence and accommodation changes may be used to elicit either constriction or dilation, while the loud sound and shock usually cause only a dilation. The third study used a change in convergence and accommodation to bring about constriction of the pupil as the UCS.

Apparatus and method.—The equipment and the booth used in Exp. I and II were also used for this study. Inducing the change in convergence and accommodation involved two bull's-eye fixation points and several mirrors. The near fixation point, which was 7 in. from *S*'s eye, was seen through a hole in a small vanity mirror, which was mounted 6 in. from the eye. The far fixation point, which was seen reflected from the face of the small mirror and from a second mirror mounted above and behind *S*, was 10 ft. from the eye. The lights transluminating the red bull's-eyes were made as dim as possible while still providing clear vision. The fixation points were connected through a set of Hunter and Eagle timers so that one went off as the other went on; the time intervals could be varied as desired. The CS was the bell used in the first study, the intensity was greatly reduced by using 4 v. AC and by moving the bell to the floor and a distance of 5 ft. from the ear. The loudness of the bell was not calibrated, but the pupil response to the bell permits comparison with the intensity used in Exp. I and IV.

In a conditioning trial, the bell began ringing .5 sec. before focus was shifted from the far to the near fixation point. The bell and the near fixation point were on together for 4 sec., and went off at the same time. The far fixation point appeared immediately and *S*'s eyes diverged to fixate on it for the 10.5 sec. between trials. Conditioning trials were presented in units of 25 with a test trial following each unit. Pictures were taken at the rate of 2 per sec. during the test trial. A 3-min. rest period was interposed between the groups of trials. A

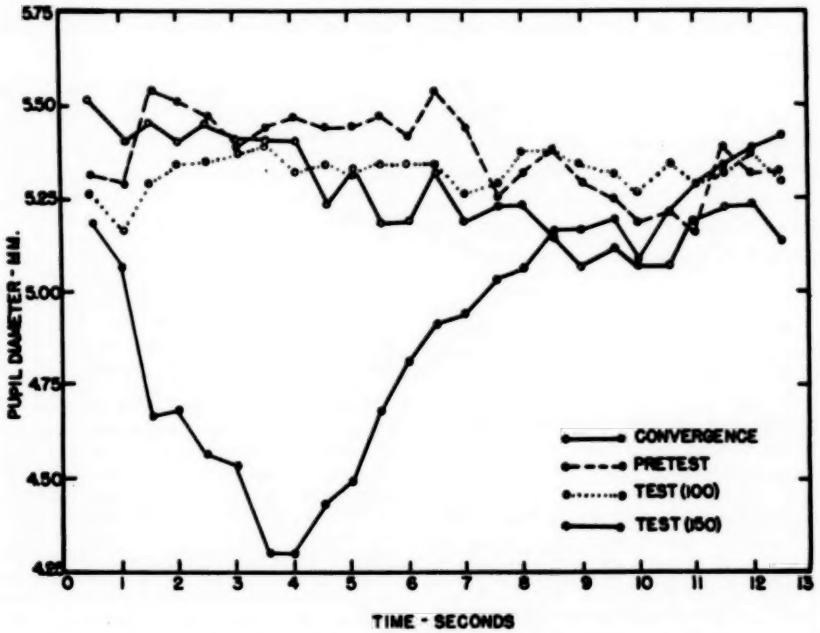

FIG. 4. The pupillary response of male *S*s to a change in convergence (UCS) and a bell (CS) and after 100 and 150 pairings of the CS and UCS.

total of 150 conditioning trials were given to each of five male *S*s of college age. Two pretests were given: one for the response to the bell alone and a second for the pupil response to the change in convergence. All trials were made in complete darkness except for the fixation lights.

Results.—The results are presented in Fig. 4. The pupil response to the bell was again a dilation, but the magnitude was only one-half that found for the stronger stimulus used in Exp. I (see Fig. 1). The response to the change in convergence from a far to a near fixation point and back was a typical pupil constriction and dilation curve (**6, 13**), but with a smaller range than can be elicited with a change in light intensity. The curve, based on the averaged results of the test trial following 100 conditioning trials, is essentially linear and parallel to the abscissa with a normal amount of fluctuation, and indicates no conditioning. The curve for the test following 150 conditioning trials is somewhat more like the pretest curve than the 100 trial test curve, but still bears no relationship to the type of curve found with the change in convergence and accommodation.

Discussion.—While the changes in convergence and accommodation with a shift in fixation points from 10 ft. to 7 in. and back to 10 ft. elicit a typical constriction-dilation pupillary response which should be adequate as UCS, no clear evidence of pupillary conditioning was obtained when the convergence change was paired with a bell as the CS for 150 conditioning trials. Since the magnitude of the pretraining response to the bell was about half that elicited in Exp. I, when the bell was approximately twice as loud, it appears that the magnitude of the pupil response is directly related to the strength of the stimulus.

Experiment IV

In this study a mild shock was used to elicit a pupillary dilation and as the UCS. The shock was paired with the bell (CS), which had an intensity equivalent to that used in the first study. The questions of the intensity of the bell and the constancy of the shock may be handled in a number of ways. The intensity of the bell may have any value, but the higher the intensity the more likely it is to arouse a pupil dilation and, up to a point, the larger will be the dilation. When using a shock as the UCS, one is faced with the question of which to maintain constant, the physical intensity of the shock or the response to the shock, since it is virtually impossible to keep both constant. In this experiment the decision was made in favor of a high bell intensity and a constant response to the shock.

Apparatus and method.—An inductorium and a galvanic skin response (GSR) recorder were added to the equipment used before. The camera was set to take two pictures per second with the presentation of stimuli, the control of the camera being governed by the timing apparatus. To insure constancy of the response to the shock, the response was monitored by connecting *S* into the galvanic skin response bridge of a Stoelting Deceptograph, Model No. 22495. While the GSR itself is not entirely constant, its variation was held to a minimum by carefully cleaning the finger used and coating it with electrode paste. The *S*s were instructed to rest the hand, immobile, on the table.

After *S* had been connected to the Deceptograph and the plated shock electrodes had been attached to the wrist of the opposite hand, the intensity of the shock was increased until *S* reported that the shock was fairly strong (enough to induce a slight twitch of the hand to which the shock electrodes were attached) but bearable. If this intensity gave a satisfactory GSR record, this was used as the initial value. If the record was not satisfactory, the intensity was increased still further until an acceptable record was obtained or *S* quit (one *S* was lost in this manner). After the initial setting, the magnitude of the response indicated on the recorder

was continually observed, and the intensity of the shock was varied to keep the response as constant as possible.

The lights in the experimental booth were left on so *S*'s pupil constricted to an average size of 4.7 mm. (4.3 mm. for males and 5.2 mm. for females), i.e., just below the midpoint of the total range of variation in *Ss* between 18 and 38 yr. old (**13**), which permits a dilation of 2 to 3 mm.

The bell, which was connected to the 8 v. AC source, began ringing .6 sec. before the onset of the shock, and rang for .8 sec. The duration of the shock was .2 sec. Fourteen seconds elapsed between the cessation of the ringing and its onset for the following trial.

A pretest for response to the bell alone was made, followed by a pretest to the shock alone. In each of these tests and in all conditioning tests, pictures were taken at the rate of 2 per sec. for a total of 14 sec. beginning 1 sec. before the onset of the bell. Following the pretest, *S* received one conditioning trial, which was immediately followed by a test for conditioning. This, in turn, was followed by two more conditioning trials and a test; four conditioning trials, a test, and a rest period. After the rest period, *S* was given eight conditioning trials, followed by a test and another rest period. Following this *S* was given 10 conditioning trials for a total of 25 trials in which the bell and shock were paired. Trial 25 was followed by a test and a rest period. After the rest period, extinction trials were presented in units of 10, followed by a test and a rest period, until *S* showed no response on the GSR for at least three consecutive trials. Two consecutive sets of test pictures were made immediately following the three trials without a GSR response, which completed the series. All rest periods lasted 3 min.

Seven male and eight female *Ss* were used in this study, and the complete experiment was run in one experimental session for each *S*.

Results.—The results for the men are presented in Fig. 5 and for the women in Fig. 6. To permit a direct comparison, these average curves are plotted on coordinates the same as those used in Exp. I, but are based on two rather than four measures per second.

Discussion.—For the men, the curves obtained in Exp. IV show the same general characteristics as those in Exp. I, when allowance is made for the 4.5 sec. bell duration in Exp. I and the .8 sec. in Exp. IV. For the women, the pretest curves for Exp. IV are a little less similar to the curves obtained in Exp. I, but still show a similar linearity which is independent of time. The later curves for the female *Ss* in this experiment appear quite different from the pretest curves, and more nearly approximate the curves obtained on the males. The curves in Fig. 5 and 6 show a considerable change in pupillary diameter with time spent in the experimental situation, the decrease in diameter being much more noticeable in the males. While the records made on the males in Exp. I contain no evidence of such a change, the final curves for the women in Exp. I indicate a pupil diameter .5 mm. smaller than that in the earlier curves. The differences between the curves of Exp. I and IV must be related to the differences between the

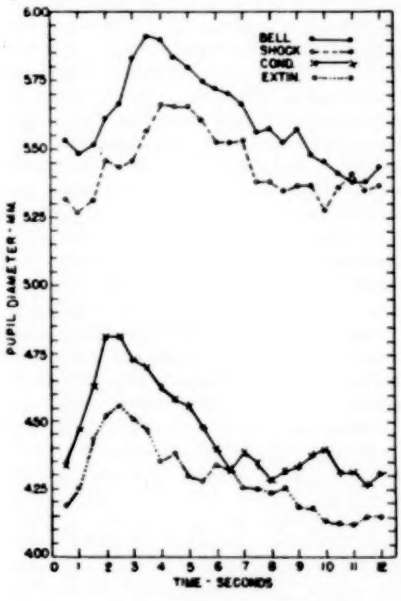

FIG. 5. The pupillary response of male *Ss* to a shock (UCS), a bell (CS) and after 25 pairings of the CS and UCS, together with the response following 20 or more unreinforced presentations of bell.

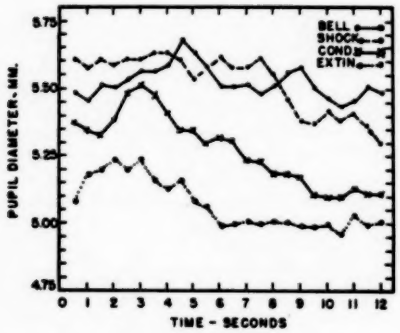

FIG. 6. The pupillary response of female *Ss* to a shock (UCS), a bell (CS) and after 25 pairings of the CS and UCS, together with the response following 10 or more unreinforced presentations of the bell.

two experimental conditions. There are three important differences: (*a*) shorter duration of the CS, (*b*) higher level of general illumination and consequently smaller pupil diameter, and (*c*) a more highly emotionally charged situation occurring in Exp. IV. As far as *a* is concerned it should not affect the pupillary diameter from one test trial to another, and thus may be eliminated from further consideration. The higher level of illumination (*b*) has the effect of increasing the range over which any emotional dilation may operate and would not influence the nature of that dilation as long as the pupil is constricted enough initially to allow the dilation to occur. The differences between the curves must be related to the emotional aspects of the experimental situations.

Lowenstein (6) has shown that strong emotional excitement causes the pupil to dilate to a larger diameter, which serves as a base line from which the pupil responds to light in the normal way. As the emotion subsides, the pupil constricts to the normal level for the given light conditions. Since the pretest curve for the female *Ss* in Exp. I, when they were not aware of the nature of the experimental situation or its onset, shows an initial rise similar to that in the curve for the males, it would appear that the women started at a normal level for the given illumination and proceeded to an emotional state which did not relax during the picture-taking span of the test trial. This hypothesis seems to be supported by the women's universal complaint about the disturbing characteristics of the bell. If it is assumed that when the general emotional state is high enough, the pupil is so dilated that it cannot expand farther to a fleeting emotional stimulus, then the women did not respond to the bell like the men did because the emotional state was too high. With further experiences, the emotional state and, along with it, the size of the pupil are likely to decrease. The final test curve for females in Exp. I does show a decrease in pupil diameter, but does not show any response to the bell itself. Apparently, Exp. I contained no particularly strong emotional stimulus for the men, and thus, they responded normally to the bell and their pupil diameters did not change with further experience.

The situation in Exp. IV was different for both men and women. Both showed heightened emotional states. For the men, it was not high enough to eliminate the response to the bell and the shock; whereas, it was that high for the women. With further experience, both groups showed a decline in the emotional level, the men showing a greater change than the women. The emotions of the female *Ss*, however, declined enough that the response to the bell became quite evident.

Both groups showed a more rapid response on later tests. On the earlier test trials, the bell had stopped before the response began, but on later trials, the response began just after the onset of the bell.

If an evaluation of conditioning is based on the males' results, there is no evidence of conditioning since none of the curves differ in any way from the original UCR to the bell. According to the logic under which the experiment was set up the two responses, to the bell and to the shock, should have been additive, and a CR should have had a

magnitude larger than that of either one alone. This possibility may be restricted by an inherent restriction on the pupil's response to an emotional stimulus, and thus dilation would not follow an increase in the intensity of the stimulus. No support for this position is available at the moment. Perhaps, a better test of conditioning would have been achieved if a less intense auditory stimulus had been used.

General Discussion

Since the repetitions of Baker's (**2**) and Hudgins' (**5**) studies did not support their findings, an explanation of this failure is warranted. Both studies were repeated exactly in every detail except two: the substitution of a camera for the telescope and kymograph used in both studies and the use of different subjects. The latter is a necessary condition and may be ruled out, since we are dealing with a physiological function which is remarkably consistent from *S* to *S* and since all experiments involved more or less randomly selected college students. It is rather inconceivable that randomly selected *S*s from similar populations should give rise to such divergent results. The different results must, then, be related to the only major difference between the studies—the method of measurement.

The use of the combination of a telescope and a kymograph opens the way to a number of sources of error. Errors will arise if the relationship between the telescope and kymograph is varied in any way: e.g., movement of the telescope, stretching of the string, string winding on top of itself, changes in temperature and humidity, movement of the kymograph. Other errors may arise because *E* cannot follow the movements of the pupil as they occur, cannot control accurately his own hand movements, etc. Further sources of error in both Baker's and Hudgins' experiments arise out of the absence of automatic controls. The task of using the telescope alone is so demanding that *E* does not have time to present the stimuli and manipulate the other apparatus. This is evident in the records given in Hudgins' study (**5**). He tried to ring the bell for 5 sec. on a test, but the records indicate that it actually rang anywhere from 4.5 to 12 sec.

In the evaluation of Baker's results, at least four factors make his results questionable. One of these is the magnitude of the CR. There is little in the way of precedent for obtaining a CR which is larger than the original UCR, yet his CR is twice as large as the UCR—3 mm. CR to 1.4 mm., the average response to the change in illumination. A second consideration is his use of a 16-power magnification in a telescope with the objective only 7½ in. from the eye. This setup would give an image of the pupil which subtends between 10° and 15° of visual angle, placing the edges of the pupil out in the parafoveal area of *E*'s eye and thus greatly increasing the difficulty of keeping the hairlines in contiguity with the edges of the pupil. Thirdly, the curves which Baker obtained are quite unlike the curve of normal pupillary response, regardless of the stimulus used. The dilation part of his convergence curve rises more steeply than the convergence part drops, an occurrence which is contrary to all published data concerning the constriction of the normal pupil. Further, there is no precedent for obtaining conditioning as rapidly as he obtained it under his experimental conditions. The measuring and recording devices used, coupled with the magnification used, would be likely to make accurate recording quite difficult and to allow factors within *E* to determine the nature of the results.

Since many, but not all, of the factors indicated above also apply to Hudgins' experimental situation, his results will be considered in terms of the findings in Exp. I and IV of this paper. To facilitate this discussion, Hudgins' Fig. 2 and 3 have been reproduced together in Fig. 7. Hudgins reported his data in two forms: (*a*) values which represent the increase or decrease in the diameter of the pupil at unspecified intervals related

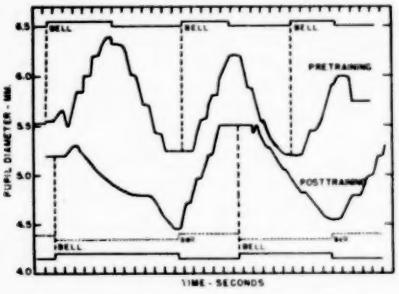

FIG. 7. Reproduction of Hudgins' Fig .2 and 3 replotted on the same coordinates.

to the sounding of the bell or the presentation of the other stimuli which he employed in his experiment, and (*b*) a set of kymograph records which are based on two different *S*s.

Without specification of the time at which the readings were made and without taking into account the change in latency which occurs with training, these values are meaningless, as may be demonstrated by an examination of the pretraining curve in Fig. 7. If the first reading is taken at the onset of the bell, the value is about 5.5 mm.; 5 sec. later it is 6.4 mm., or a dilation of .9 mm. which is greater than his average value of .57 mm. However, if one takes the reading 5 sec. later, the value is now 5.25 mm. or a constriction of .25 mm., as compared with the .31 mm. average constriction he obtained after training. Thus before training, dealing only with the response of the pupil to the bell, one can get the same values as he obobtained with training, provided one chooses the right time to make a reading. Nowhere does Hudgins indicate the time relationships involved in the measurements. The significance of time of reading, coupled with the decrease in reaction latency accompanying training, requires rigid control of the time elapsing between onset of the stimulus and the measurement of the response, and coordination of this time interval with amount of training in order to obtain meaningful results. The lack of these controls in Hudgins' experiment makes the validity of his results quite questionable.

Together with the data examined above, Hudgins presented actual kymographic records of two *S*s. The curves for one *S*, which is purported to indicate a conditioned constriction response to the bell, are reproduced in a slightly modified form in Fig. 7. Hudgins indicates that the method of testing for the conditioning effect of the bell was in every respect identical with the method used in the initial measurements. Measurements for the initial effects of the bell were made in the following manner, "The light was turned on and the pupil was measured as in the other tests. Then the experimenter rang the bell for five (5) sec. and again measured the pupil. Sometimes several measurements were made without extinguishing the light" (**5**, pp. 9–10). Since Hudgins stressed the duration of the ringing of the bell, but did not specify the amount of time elapsing between the bell and the measurement, it may be assumed that this interval was not controlled. Hudgins consistently made the point that 10 sec. elapsed between a measurement after the bell and the next ringing of the bell (**5**, p. 9). Applying this sequence to the kymographic recording, one would expect to find that the bell rang for 5 sec. and was off for 10 sec. in any one test cycle, but in Fig. 7, the time intervals are 6.5, 5, and 4.5 sec. for the bell-on and 7 and 6 sec. for the bell-off. These times for the bell-on are probably as close to 5-sec. intervals as could be obtained without a mechanical timer. Examination of the posttraining curve of Fig. 7 discloses an interesting discrepancy. In Hudgins' words, "This is a complete reversal of the response which appeared in the record taken prior to conditioning" (**5**, p. 17). It is a complete reversal in more ways than one.

Contrary to his experimental conditions and the conditions used in the pretraining test, the bell was not run for 5-sec. but for 12- and 9-sec. periods, and the off period was not approximately 10 sec. but was 6 sec. If it happened that

the direction of the signal marker was reversed in this record, then, actually the bell was on for 6 sec. and off for 12 and 9 sec., a condition suggested by the dotted line in Fig. 7, and the records would then be comparable to the pretraining curve. Reversal of direction certainly brings this record into closer agreement with the stipulated experimental conditions. If the reversal is assumed, his constriction curves become dilation curves of a form which closely resembles that of the curves obtained in Exp. IV of these studies. Actually, the dilation curves for his pretraining tests are much too symmetrical to be actual recordings of the pupil response, and the post-training curves more nearly approximate the curves for the dilation response to the bell. In this connection it is interesting to note that the magnitude of the conditioned constriction response of this *S*, between .75 and 1 mm., exceeds that of any of the *S*s measured in the first part of his experiment by almost one *SD*, but falls within the range of the dilation response to the bell.

This simple reversal does not explain the 1-mm. decrease in the size of the pupil. This difference is relative, and may be caused by changes in apparatus between preliminary and final testing, or may represent a phenomenon similar to that shown in Exp. IV, where the size of the pupil decreased with time spent in the experimental situation. Since Hudgins' bell was quite loud and disturbing, conditions were present which would allow either or both of these explanations.

Since neither Exp. III nor Exp. IV, along with I and II, gave any evidence of conditioning, they do not support the conclusion that it is possible to condition the pupil. The analyses of Baker's and Hudgins' experiments presented above make it questionable whether any experiments have shown conditioning of the pupil.

Summary

Exact repetitions of Baker's and Hudgins' experiments except for the substitution of an infrared motion picture camera for the telescope and kymograph used in the original studies have failed to confirm conditioning of the pupil. Attempts to condition the pupil to respond to a bell when convergence and accommodation or shock were used as the UCS also failed to give evidence of conditioning.

An analysis of Baker's experiment indicates that the visual angle subtended by *S*'s pupil in his telescope made accurate measurement extremely difficult, and could have allowed set and attitudes within the E to influence the measurements. Analysis of Hudgins' experiment shows that his evidence for conditioning was obtained under conditions which were the exact reverse of his stipulated experimental conditions. Further, when these conditions are brought into agreement with the given experimental conditions, the results agree with those obtained when the experiment was repeated. In these circumstances, the conditioned constriction turns out to be a dilation response to the bell similar to that obtained before conditioning and to that obtained in this repeat study.

When the results obtained here and in an earlier study (**14**) are integrated, it appears that different types of conditioned stimuli (bell and pure tone), different intensities of the same type of stimulus, different durations of the stimulus, different intervals between the onset of the conditioned and unconditioned stimuli, and different types of unconditioned stimuli are all incapable of eliciting pupillary conditioning in either sex.

References

1. Baker, L. E. The influence of subliminal stimuli upon verbal behavior. *J. exp. Psychol.*, 1937, **20,** 84–100.
2. Baker, L. E. The pupillary response conditioned to subliminal stimuli. *Psychol. Monogr.*, 1938, **50,** No. 3 (Whole No. 223).
3. Crasilneck, H. B., & McCranie, E. J. On the conditioning of the pupillary reflex. *J. Psychol.*, 1956, **42,** 23–27.
4. Hilgard, E. R., Dutton, C. E., & Helmick, J. S. Attempted pupillary conditioning at four stimulus intervals. *J. exp. Psychol.*, 1949, **39,** 683–689.
5. Hudgins, C. V. Conditioning and the voluntary control of the pupillary light reflex. *J. gen. Psychol.*, 1933, **8,** 3–51.
6. Lowenstein, O. Pupillary reflex shapes and topical clinical diagnosis. *Neurology, Minn.*, 1955, **5,** 631–644.
7. Seashore, C. E. An audiometer. *Univer. Iowa Stud. Psychol.*, 1899, **2,** 158–163.

8. Steckle, L. C. Two additional attempts to condition the pupillary reflex. *J. gen. Psychol.*, 1936, **15,** 369–377.
9. Steckle, L. C., & Renshaw, S. An investigation of the conditioned iridic reflex. *J. gen. Psychol.*, 1934, **11,** 3–23.
10. Stern, F. An investigation of pupillary conditioning. Unpublished doctor's dissertation, Univer. of Washington, 1948.
11. Wedell, C. H., Taylor, F. V., & Skolnick, A. An attempt to condition the pupillary response. *J. exp. Psychol.*, 1940, **27,** 517–531.
12. Young, F. A., & Biersdorf, W. R. An apparatus for taking rapid photographs of the pupil in visual darkness. *Amer. J. Psychol.*, 1952, **65,** 617–618.
13. Young, F. A., & Biersdorf, W. R. Pupillary contraction and dilation in light and darkness. *J. comp. physiol. Psychol.*, 1954, **47,** 264–268.
14. Young, F. A. An attempt to obtain pupillary conditioning with infrared photography. *J. exp. Psychol.*, 1954, **48,** 62–68.
15. Zabarenko, R. N. The conditioning of the pupillary light reflex. Unpublished master's thesis, Univer. of Pittsburg, 1939.

(Received February 8, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

"BACKWARD" ASSOCIATIONS IN TRANSFER AND LEARNING[1]

BENNET B. MURDOCK, JR.

University of Vermont

In a previous experiment (**3**) it was shown that in verbal paired-associate learning "backward" learning does occur. That is, when *S* is learning A—B he is at the same time learning B—A. This formation of backward associations may be a trivial and insignificant part of the learning process, or it may be important. The present study is an attempt to obtain further information about backward association by testing two hypotheses about their effects on transfer and learning.

The first hypothesis is that the backward associations developed in the course of learning one list will transfer to a second list. It is generally considered that A—B, A—C produces negative transfer; in learning the second list the B responses are evoked by the A stimuli and interfere with the acquisition of the new C responses. By the same type of reasoning, if the backward associations transfer, then A—B (B—A), B—C should also produce negative transfer.

The second hypothesis to be tested is that previous learning which interferes with the backward associations formed in learning a given list will retard the learning of the list. This can be tested by the transfer paradigm A—B, C—A. In learning C—A in the second list *S* is presumably also learning A—C. The forward associations from the first list would interfere with the formation of the backward A—C associations on the second list, and if this interference affects learning negative transfer should again occur.

A previous study by Harcum (**2**) provides some evidence for both these hypotheses. He found negative transfer with A—B, B—C and with A—B, C—A; in both cases the learning of the second list was slower than the learning of a control list. However, from the report it is not clear whether possible list-differences were controlled. Also, in one case (A—B, C—A) the transfer was not statistically significant and in the other (A—B, B—C) the variance was heterogeneous. In view of the uncertainty of these results, it was felt that further investigation was in order.

Procedure

There were three groups in the present experiment; one for A—B, B—C, one for A—B, C—A, and one for A—B, A—C. The third group, the standard negative transfer paradigm, was included to evaluate the magnitude of the transfer effects found in the first two groups. There were 60 *Ss*, male and female students enrolled in undergraduate psychology courses. The *Ss* were randomly assigned to groups, with 20 in each condition.

The design was essentially the same one that was used in the previous study of backward learning (**3**), and is described elsewhere as Design IV with counterbalancing (**4**). For all *Ss* the first or A—B list was 6 pairs long, and the second list 12 pairs long. Six of these 12 were control pairs (C pairs) which can be labelled C—D; i.e., pairs with new stimuli and responses. The remaining six were experimental pairs (E pairs), either B—C, C—A, or A—C. The E and C pairs were randomly intermixed within the list, so in learning one

[1] This research was in part supported by a grant, G-2590, from the National Science Foundation. Mr. James Barnard tested half of the *Ss* and helped in the analysis of the results.

list *S* learned both types simultaneously. In this way each *S* served as his own control, and warm-up and learning-how-to-learn were controlled. Transfer was the difference between the learning of the E and C pairs. Counterbalancing was used in order to control for possible inequalities in ease of learning between E and C pairs; E and C pairs were reversed to form C and E pairs, respectively, for half the *Ss*.

For the two lists a total of 30 syllables were needed (12 for the A—B pairs, 12 for the C—D pairs, and 6 for the C terms in the 6 E pairs). These 30 were selected from the list of Glaze (**1**) 100% association-value nonsense syllables. Duplication of letters and letter combinations was minimized as far as possible. The individual syllables were randomly paired to form the necessary groups (A—B, A—C, B—C, C—A, and C—D), although any obvious similarities between the two members of a pair were prohibited. Thus, in this experiment both inter-list and intra-list similarity were relatively low.

As an additional precaution, for each of the three groups the pairing of individual syllables was done twice, and each pairing was independent of the other (except that they both started with the same 30 syllables). This then provided two complete replications of the experiment, each using the same 30 syllables but different S—R combinations. Half the *Ss* were assigned to each pairing, and within each pairing counterbalancing was used.

All *Ss* first learned a practice list, which was nine pairs long and consisted of Glaze 53% association-value syllables. The practice list was learned to a criterion of three correct anticipations on one trial. The first list was learned to a criterion of three consecutive perfect trials, and the second list was learned until each of the 12 pairs had been correctly anticipated at least once, although not necessarily on the same trial.

For all lists the presentation rate was 3 sec. (1.5 sec. for S; 1.5 sec. for S—R). Anticipation, pronunciation, noncorrection, and a varied order of presentation were used. There was no inter-trial interval; the first pair of one trial followed immediately the last pair of the preceding trial.

Results

In List 2, each *S* learned six E and six C pairs. For each *S* the transfer was measured by the formula, T = E − C/E + C. The general advantages of this formula are discussed elsewhere (**4**), but essentially they are that positive and negative transfer are symmetrical and both have an upper limit of 100%.

For the 20 *Ss* in the A—B, B—C group the mean transfer was −11.2% (SD = 12.3); for the A—B, C—A group, −12.2% (SD = 12.2); and for the A—B, A—C group, −10.2% (SD = 19.2). Each of these values differed significantly from zero (t's[2] of 3.96, 4.36, and 2.31, significant at the .01, .01, and .05 levels, respectively). Thus, all three groups showed negative transfer.

An analysis of variance showed that the three groups did not differ significantly ($F < 1.00$). Also, the two pairings gave almost identical results ($F < 1.00$). This latter finding suggests that the results do not depend upon the particular pairing of individual stimulus and response terms.

For all *Ss*, a mean of 19.6 trials was required to learn List 2 to criterion. The means for the three groups were 17.8, 21.4, and 19.6 trials, and these values did not differ significantly. For all *Ss*, the correlation between the number of trials to learn List 2 to criterion and the percentage transfer was practically zero (r = .02), thus showing no relationship between speed of learning and amount of transfer.

To complement percentage transfer as a measure of transfer a transfer score (T = E − C) was also computed. For the three groups the mean transfer scores were −2.6, −2.1, and −1.8, respectively. Thus,

[2] Since the 60 transfer scores formed an almost perfectly normal distribution (M = −11.2, SD = 14.9), and since the measure of transfer differs somewhat from the usual percentage, it was felt justifiable to use t tests to evaluate the statistical significance. Actually, the experimental hypotheses would still be confirmed even if the least powerful of the non-parametric statistics, the sign tests, had been used.

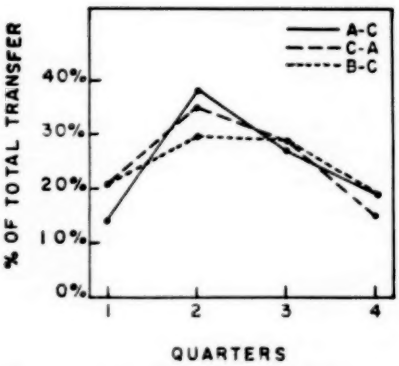

FIG. 1. Percentage of total negative transfer occurring in each quarter of learning.

on the average, C pairs were correctly anticipated approximately two more times than E pairs.

The measure of learning used (total number of correct anticipations of each pair) reflects performance over the entire learning period. To determine at what stage or stages of learning the negative transfer occurred a trial-by-trial analysis of the transfer score was made. Vincent curves for the three groups are shown in Fig. 1; for each group the curve shows what percentage of the total (negative) transfer occurred in each quarter. Since the three groups were almost identical they were combined to test for transfer as a function of stage of practice. There was a significant difference among the quarters ($F = 4.42$, $P < .01$), and the second and third quarters combined showed significantly ($P < .01$) more negative transfer than the first and fourth quarters combined. Thus, the transfer was least at the beginning and end of the second task learning.

On the average, 8.5 trials were required to learn the practice list to criterion and 24.8 trials were required to learn List 1 to criterion. In neither case did the difference among groups approach significance, thus indicating that the groups were probably reasonably equivalent in learning ability.

DISCUSSION

The results clearly support both hypotheses. Thus, it would seem that under the present conditions backward associations do transfer, and that previous learning which interferes with the formation of backward associations in a new learning situation will retard the learning. Also, in view of the fact that the A—B, A—C condition showed if anything less negative transfer than the other two conditions it would almost seem that backward associations may have as great an inhibitory potential as forward associations.

The results of this experiment can also be interpreted in quite a different way. According to the Müller-Schumann law of associative inhibition, "Once any two items have been associated, it becomes more difficult to associate either of them with a third term" (**5**, p. 525). The results obtained are exactly those that would have been predicted from this law. However, as Osgood (**5**) has pointed out, this law is not so satisfactory in explaining transfer results when higher levels of inter-list similarity are used.

The results of this experiment are not incompatible with the results obtained from studies of serial learning. Serial learning in a forward order does occur; this is attested to by everything from the fact that the alphabet is notoriously difficult to recite backward to the fact that following serial learning forward associations are almost always considerably stronger than backward associations. However, the method of serial learning very definitely imposes a set to learn in the forward direction since the items are always presented in the same order. In paired-associate learning, varying the order of presentation is done in an attempt to prevent such

a set from becoming established; the order is essentially random rather than constant. It is, then, under the paired-associate method that the present experiment has shown that backward associations both transfer and affect learning.

Summary

An attempt was made to determine if the backward associations formed in learning one verbal paired-associate list transfer to a second list, and if interference with the formation of backward associations retards learning. The results indicated that the backward associations do transfer, and that interference with backward associations does retard learning. The results also suggest that, in terms of magnitude, the effects of backward associations may be as great as the effects of forward associations.

References

1. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–269.
2. Harcum, E. R. Verbal transfer of overlearned forward and backward associations. *Amer. J. Psychol.*, 1953, **66,** 622–625.
3. Murdock, B. B., Jr. "Backward" learning in paired associates. *J. exp. Psychol.*, 1956, **51,** 213–215.
4. Murdock, B. B., Jr. Transfer designs and formulas. *Psychol. Bull.*, 1957, **54,** 313–326.
5. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

(Received January 15, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

PROBABILITY LEARNING UNDER EQUIVALENT DATA COLLECTION METHODS [1]

S. S. KOMORITA [2]

University of Michigan

The majority of probability learning studies dealing with the non-contingent procedure have been restricted to situations involving two stimuli where one and only one stimulus event occurs on each trial. Recently, however, experiments have been conducted where *S*s were asked to predict from more than two alternatives (**1**, **4**, **8**). These studies suggest certain extensions in theory and methodology appropriate to multiple-choice behavior. Accordingly, this paper is concerned with asymptotic response probabilities in probability learning situations under the following conditions: (*a*) non-contingent procedure where stimulus feedback is independent of *S*s' responses, (*b*) three stimulus alternatives, (*c*) more than one stimulus event may occur on each trial, and (*d*) *S*s may be asked to make more than one prediction per trial. Under these conditions of increased complexity certain interesting methods of data collection are possible.

Coombs (**3**) has recently proposed a theory of data in which methods of collecting data have been ordered and classified. One aspect of this classification is that given n stimuli to be judged or evaluated, the stimuli may be presented k at a time, ($0 < k \leq n$) and *S*s may be asked to *Pick* m or *Reject* m of the k stimuli ($0 < m < k$). The interesting point here is that from a logical position asking *S*s to Pick m-1 is equivalent to asking them to Reject 1 (or asking them to Pick 1 and Reject m-1).

In order to make predictions in the complex situations considered in this study a mathematical model was constructed based upon the following postulates: (*a*) Whenever *S* evaluates, compares, or judges a set of stimuli, the basic psychological process is a discrimination of paired subsets of the stimuli. (*b*) People "attend to" or "notice" particular subsets of the stimuli. Subsets of the stimuli, therefore, may be ordered on the basis of their "attention value." (*c*) The order in which people evaluate subsets of the stimuli is a function of the "attention value" of the stimuli. That is, stimuli with the greatest "attention value" are evaluated first and stimuli with the least "attention value" are evaluated last.

It should be pointed out that the general model presented here is not limited to probability learning situations but may also be applicable in other situations such as the psychophysical area or the general area of scaling. Whereas models constructed by Bush and Mosteller (**2**) and by Estes (**5**, **6**) represent statistical theories of learning, the model presented in this paper may be more appropriately described as one which represents a theory of data collection. Consequently, the axioms of this

[1] Based upon a dissertation submitted to the Psychology Department, University of Michigan in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to thank Dr. David Birch for his invaluable advice and suggestions in carrying out this study.

[2] Now at Vanderbilt University.

model should not be interpreted as axioms of learning but axioms concerning the psychological processes underlying *S*'s choice when asked to evaluate a set of stimuli.

In a probability learning situation, if it is assumed that the only distinguishing characteristic among the stimuli is their frequency of occurrence, then the "attention values" of the stimuli are determined by their frequencies. Under these conditions the stimuli may be ordered from greatest to least according to their "attention values." On the basis of these postulates, models for the Pick *m* and Reject *m* methods of data collection can be constructed. The notation to be used here and in subsequent presentations will be as follows. Let A, B, C, $\cdots$, N denote stimuli with probabilities of occurrence p_A, p_B, p_C, $\cdots$, p_N respectively, such that $p_A > p_B > p_C > \cdots > p_N$.

The Pick m *procedure.*—Two models for the Pick *m* procedure are the conditional probability model and the pattern model. According to the conditional probability model, the asymptotic response probability for the ith choice is assumed to be a function of the previous choice. Letting subscripts refer to the order of the choices, the probability that stimulus A is predicted as a second choice is given by

$$\phi(A_2) = \phi(A/B_1)\,\phi(B_1) + \phi(A/C_1)\,\phi(C_1) + \cdots + \phi(A/N_1)\,\phi(N_1), \quad (1)$$

where N denotes the least frequent stimulus. This means that Stimulus A will be predicted as second choice a given proportion of times that each of the others are predicted as a first choice. Similarly,

$$\phi(J_2) = \phi(J/A_1)\,\phi(A_1) + \phi(J/B_1)\,\phi(B_1) + \cdots + \phi(J/N_1)\,\phi(N_1). \quad (2)$$

Similar equations may be derived for the ith choice; the derivations for the second choice were given here for simplicity of illustration. The asymptotic response probability for stimulus J is then given by

$$\phi(J) = \sum_{i=1}^{m} \phi(J_i). \quad (3)$$

The pattern model, on the other hand, may be described as one in which *S* predicts in terms of the frequency with which certain stimuli occur together. That is, he responds to the patterns or combinations of stimuli occurring together.

The Reject m *procedure.*—In the Reject *m* procedure *S* is asked to predict which stimuli will *not* occur; i.e., he is asked to Reject *m* rather than Pick *m*. It follows from Postulates b and c that *S* "attends to" the least frequent stimulus rather than the most frequent. The stimuli may then be redefined in terms of their probabilities of non-occurrence and this reduces to the Pick *m* case with the order of comparison reversed.

The matching model.—The Pick 1 case is a special case of the other models. One of the many possible functions for the Pick 1 case is the matching model. The matching model may be defined as follows:

$$\phi(J) = \frac{p'_j}{\Sigma p'_i}, \quad (4)$$

where p' refers to psychological or subjective probability.

The matching model predicts that *S* will respond in proportion to the subjective probabilities of the stimuli. The psychological interpretation is that *S* makes a paired comparison of each stimulus with all of the stimuli combined. Since subjective probabilities are as yet an undetermined function of the objective probabilities, the present study will assume that they are equal. With this assumption Equation 4 is equivalent to equations derived by Bush and Mosteller and by Estes (**2**, **5**, **6**).

In view of these considerations the purpose of the experiment to be reported may be summarized as follows: (*a*) to determine to what extent different methods of data collection, equivalent from a logical position, yield equivalent empirical results; (*b*) to determine some of the situations to which matching behavior, as defined by the matching model, can be generalized; and (*c*) to test some of the predictions made by the model presented in this paper.

Method

Subjects.—The *S*s were 90 male undergraduates at the University of Michigan. All were

enrolled in introductory psychology and their participation as *Ss* fulfilled part of their requirements for the course.

Apparatus.—Three lights were mounted horizontally and equally spaced on a rectangular panel. In all cases the most frequent light was on the left and the least frequent light on the right. The *S* had three telegraph keys before him, each key corresponding to one of the lights. The apparatus also included a programmer which presented the stimuli from punched tape, and electrical counters which recorded the responses of *Ss*. With the use of this apparatus each trial consisted of: (*a*) a bell as a ready signal, (*b*) a 4.8-sec. period during which *S* was to make his response(s), (*c*) a 1.6-sec. waiting period during which time no responses would be recorded by the apparatus, and (*d*) a 1.6-sec. period during which the light(s) came on.

Experimental conditions.—Two random probability distributions were used. One was a .60 : .30 : .10 series and the other a .90 : .70 : .40 series; they will hereafter be referred to as Series A and B, respectively. Both series were randomized so that blocks of 50 trials conformed to these distributions. Series A was randomized such that on each trial one of the lights came on. Series B was randomized such that on each trial two of the lights came on; i.e., over each block of 50 trials the .90 and .70 stimuli occurred together on 60% of the trials; the .90 and .40 stimuli occurred together on 30% of the trials; and the .70 and .40 stimuli occurred together on 10% of the trials.

Each series consisted of six practice trials and five blocks of 50 trials. After each block of trials *Ss* were given a 10-sec. rest period during which time their responses were recorded from the counters by *E*.

Three methods of data collection were used: Pick 1, Reject 1, and Pick 2. Each of these methods was factorially combined with the two probability distributions so that there were 15 *Ss* in each of six experimental groups. These six groups will be referred to as Groups P1A, P1B, R1A, R1B, P2A, and P2B, where the first letter refers to Pick or Reject, and the last letter refers to Series A or B.

The setting.—The experiment took place in a small room with curtains screening *Ss* from the apparatus and *E*. In the Pick 1 and Reject 1 groups *Ss* were scheduled in pairs. However, because of scheduling difficulties and since some *Ss* did not keep their appointments, some of them were run individually. A curtain separated those *Ss* run in pairs. Both of the Pick 2 groups were run individually because of certain limitations of the apparatus.

The instructions given to these six groups were modifications of the instructions used by Estes and Straughan (**7**). The three groups which were presented Series A were instructed that one of the lamps would light on each trial, and the three groups which were presented Series B were instructed that two lamps would light on each trial. The two Pick 1 groups were instructed to guess which of the three lamps would light on each trial, while the two Reject 1 groups were instructed to guess on each trial, which of the three lamps would *not* light on each trial. The two Pick 2 groups were instructed to make two guesses attempting to predict which one (or two) of the lamps would light on each trial.

Results and Discussion

Tests for asymptotic performance.—The response measure used throughout the analyses was the frequency of responses to the most frequent alternative on five blocks of 50 trials. For a measure of asymptotic performance, frequency of responses on the last block of 50 trials was arbitrarily selected. For each of the six experimental groups *t* tests were performed to determine if responses on the last block differed from the responses on the previous block of trials. This analysis resulted in differences significant at the 5% level for Groups P1B and P2B. It should be pointed out, however, that finding no significant differences does not imply an asymptotic level; this is a necessary but not a sufficient condition. Accordingly, any generalizations about asymptotic performance must be interpreted with caution.

Comparisons of equivalent data collection methods.—The design of the study permitted six comparisons of logically equivalent data collection methods. These six comparisons, however, are not independent of one another; i.e., the six comparisons may be divided into two sets of three paired comparisons each. One set consists of the following paired comparisons: Groups P1A-R1B, P1A-

P2B, and R1B-P2B. The other set consists of the following paired comparisons: Groups P1B-R1A, P1B-P2A, and R1A-P2A. Figures 1 and 2 summarize the data for these two sets of three groups showing the probability of responses to the most frequent alternative over the five blocks of 50 trials.

An analysis of variance showed that only one of these six comparisons was significantly different. This was the comparison between Groups P1A and P2B which resulted in a difference significant at the 5% level. This comparison may be interpreted as a test of whether people respond to the pattern of double stimulus occurrences as they do to single stimulus occurrences; i.e., P2B is logically equivalent to Pick 1 of 60 : 30 : 10 where the stimuli are defined to be the patterns of two lights occurring together. Under the conditions of the present study, people did not respond to single stimulus occurrences as they did to patterns of double stimulus occurrences although the probabilities of the stimuli were identical.

One possible interpretation of this difference is that the two groups have the same asymptote but different rates of learning. Note in Fig. 1 that the difference between the two groups is almost zero initially and then becomes larger over blocks of trials. For Group P2B the patterns or pairs of lights are more similar to each other than the single lights for Group P1A. It might be assumed, therefore, that more generalization might occur for Group P2B. Hence, it is plausible to expect that Group P2B would demonstrate a slower rate of learning. This interpretation is further supported by the fact that there was reason to doubt that Group P2B had reached an asymptote.

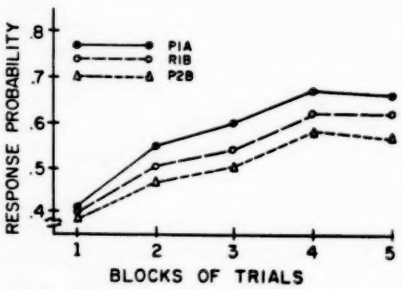

FIG. 1. Probabilities of response to most frequent alternative for first set of three equivalent data collection methods groups.

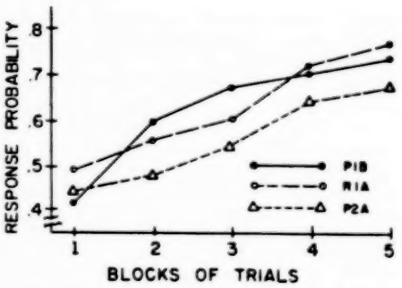

FIG. 2. Probabilities of response to most frequent alternative for second set of three equivalent data collection methods groups.

Test of the matching model.—Table 1 shows the mean frequency of responses to the most frequent alternative on the last block of 50 trials for each of the six experimental groups. In Table 1, P2B refers to the matching model when applied to the conditional probability model, while P2B(P) refers to the matching model with respect to the patterns of stimulus occurrences.

Table 1 indicates that the predictions of the matching model deviated significantly from the observed values for the P1B and R1A situations. One possible interpretation of these differences is that the matching model is

TABLE 1

TEST OF THE MATCHING MODEL: COMPARISON OF PREDICTED AND OBSERVED MEANS

Cond.	$M_{obs.}$	$M_{pred.}$	Diff.	D (.05)*
P1A	33.80	30.00	3.80	4.12
P1B	37.13	22.50	14.63	5.51
R1A	39.07	22.50	16.57	3.61
R1B	32.27	30.00	2.27	5.00
P2A	44.27	46.00	1.73	2.48
P2B	42.20	40.00	2.20	1.24
P2B(P)	29.47	30.00	0.53	1.91

Note.—Each observed mean represents mean frequency of responses to the most frequent alternative on the last block of 50 trials.
* Difference required for rejection of the matching model at the 5% level.

not applicable when the probabilities of the stimuli do not sum to unity.

With respect to the P2B situation, although the matching model used in conjunction with the conditional probability model deviated significantly from the observed values, the matching model used in conjunction with the pattern model gave an excellent fit of the data, as shown by the predictions for P2B(P). This seems to indicate that the matching model is not applicable in predicting probability of responses to the most frequent single stimulus. However, if the stimuli are redefined to be the patterns or pairs of lights occurring together, then the matching model seems to be applicable for the P2B situation. These differences, however, do not necessarily imply the rejection of the general model presented in this paper. Specific models other than the matching model, for example, may result in accurate predictions of these situations.

SUMMARY

The two main problems of this study were to determine some of the situations to which matching behavior could be generalized, and to determine if different methods of data collection equivalent from a logical standpoint yield equivalent empirical results. With respect to the first problem, the situation was extended by the use of a three-choice situation, and by the occurrence of more than one stimulus event on each trial. With respect to the second problem *S*s were asked to (*a*) pick one of the lights which they thought would come on, (*b*) pick two of the lights which they thought would come on, and (*c*) pick one of the lights which they thought would not come on. Six groups of 15 *S*s differed with respect to the instructions given to them and the stimulus series which were presented.

A general mathematical model was presented which made predictions in the complex situations considered in this study. A specific model, defined as the matching model, was used in conjunction with the general model and an attempt was made to determine how well this matching model would predict the data. Matching behavior, as defined by the matching model, was found in four of the six situations considered in this study. The results of the other two situations seem to suggest that the matching model may not be applicable when the probabilities of the stimuli do not sum to unity.

The results indicate that under the conditions of the present study, equivalent methods of data collection yield equivalent results. The one exception to this generalization was the finding that people do not seem to respond to single stimulus occurrences as they do to patterns of stimulus occurrences even though the probabilities of the stimulus events are identical.

REFERENCES

1. ANDERSON, N. H. Effect of double stimulus events in a verbal conditioning situation. *Amer. Psychol.*, 1955, **10**, 466.
2. BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.
3. COOMBS, C. H. Theory and methods of social measurement. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953.
4. DETAMBEL, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.*, 1955, **49**, 97–104.
5. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

6. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
7. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
8. Neimark, E. D. Effects of type of non-reinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

(Received January 17, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

EXPECTATION AND RESISTANCE TO EXTINCTION OF A LEVER-PULLING RESPONSE AS A FUNCTION OF PERCENTAGE OF REINFORCEMENT AND NUMBER OF ACQUISITION TRIALS[1]

DONALD J. LEWIS[2] AND CARL P. DUNCAN

Northwestern University

A modified slot machine has been found to be a satisfactory piece of apparatus to study the effects of partial reinforcement and other variables upon the behavior of human *Ss* (**6, 7**). With this apparatus two principle response measures have been taken. One has been a measure of behavior persistence—the logarithm of the number of plays to quitting (extinction). The other has been a measure of *S*'s expressed expectancy of winning or losing on the next play. It was found that the percentage of reinforcement is inversely related to mean log. plays to extinction, and that the amount of reinforcement is directly related to the same response measure. The expectancy response measure was found to be a direct function of percentage but unrelated to the amount of reinforcement. It was also found that the loss of expectancies during extinction was an almost linear function of the percentage of reinforcement. Because of this almost linear relationship, it was guessed that the expectancies might be not only a dependent variable, but they might also be an index of an intervening variable determining resistance to extinction. The present study attempts to bring data to bear on this point and to extend the findings by varying the number of acquisition trials along with the percentage of reinforcement.

METHOD

Apparatus.—The apparatus is described in detail in a previous report (**2**). It is basically an electrical slot machine which *S* plays by inserting a chromium-plated disk into a slot at the top of the machine. He then pushes four buttons out of a matrix of 16, after which he pulls a lever. The machine whirs, makes other attractive noises, and a disk drops into the payoff tray if it is a payoff trial. The machine pays off on particular plays which are determined in advance by *E*. The *S* has a very large supply of disks and can, for all practical purposes, play as long as he wishes.

Design.—One major variable manipulated in this experiment was the number of acquisition trials—3, 6, 12, and 21 acquisition trials were used, and each of the acquisition trial groups was combined factorially with three percentages of reward; 33%, 67%, and 100%, making a 12-group experiment. An *S* was paid 10¢ on each winning trial, and the last play of all acquisition series was always a winning play. The other winning plays were irregularly distributed through the acquisition series. After the appropriate number of acquisition trials, the extinction series began, and no *S* ever won again. They were allowed to play as long as they wanted, however, and the total number of plays was automatically cumulated by the machine.

In addition to the total number of responses a record was taken of *S*'s stated expectation of winning or not winning on the next trial. Secured to the front of the machine was a line rating scale with the numbers 1 through 6 placed at equal intervals on the line. Above these numbers was the title, "Expectation of winning or not winning." Corresponding to

[1] This research was supported in large part by a grant from the National Science Foundation. The authors wish to thank Dr. Marie Duncan for her help with the statistical analyses and Mrs. Judith Dinner for her help in running *Ss*.

[2] Now at Louisiana State University.

the numerical scale points from 1 through 6 were the descriptions: "fairly sure of not winning," "mildly sure of not winning," "slightly sure of not winning," "slightly sure of winning," "mildly sure of winning," and "fairly sure of winning." On each trial, before he pulled the lever, *S* called out the numerical value corresponding to his level of expectancy. The *E* sat to one side of the room and recorded the expectancy stated by *S* on each trial.

Instructions.—The instructions were identical with those of a previous experiment (**7**). The *S*'s were told how to operate the machine, and they were instructed how much money they would receive for each disk won. They were to keep the disks they won separate from those they were playing with and to turn in the former at the close of the experiment for the appropriate amount of money. The *S*s were also told that they could play as long as they liked and that when they decided to quit to exchange the disks they had won for money. After *S* had decided to quit, he was paid off with coins, given a plausible, but false, explanation of the machine, and told not to tell any other person how the machine operated or how much he had won.

Subjects.—The *S*s were drawn from introductory psychology courses at Northwestern University and were assigned to the various experimental groups in turn as they appeared in the laboratory. Seventeen *S*s were assigned to each of the 12 experimental groups making a total of 204 *S*s.

Results

Number of plays to extinction.—The number of extinction plays was determined by subtracting the appropriate number of acquisition plays—3, 6, 12, or 21—from the total number of plays. A total of 11 *S*s, however, did not complete the full number of acquisition plays. Such *S*s were considered to have 0 extinction plays. A transformation of all scores plus one to common logarithms resulted in a satisfactory normalization, and an analysis of variance was performed on these log scores.

The results of this analysis showed that the number of acquisition trials had a significant effect upon the log. trials to extinction. The F was 3.72, significant at the .05 level for 3 and 192 *df*. The percentage of reinforcement was not significant ($F = 1.41$ for 2 and 192 *df*), nor was there a significant interaction ($F = .34$ for 6 and 192 *df*). A graph of these data (Fig. 1B) shows that the larger the number of acquisition plays the fewer the number of extinction plays, except that there appears to be no difference between 3 and 6 acquisition plays.

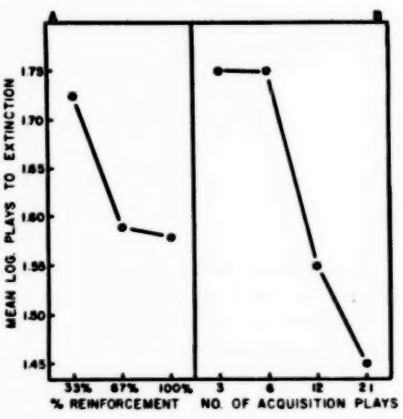

FIG. 1. (A) The relationship between mean log. trials to extinction and percentage of reinforcement. (B) The relationship between mean log. plays to extinction and the number of acquisition plays.

Figure 1 also shows the effect that percentage of reinforcement had on the number of extinction plays. Although not significant, there is a general trend indicating more plays to extinction for the smaller percentages of extinction. This trend is markedly similar to the one obtained from a previous experiment (**7**) in which percentages less than 33 were used and in which the differences were significant. For this situation it would seem that the extremes of the percentage range should be sampled in order to obtain a significant partial reinforcement effect.

Acquisition expectancies.—Analysis of acquisition expectancies during the later acquisition trials was complicated by the fact that 11 *Ss* did not finish their designated acquisition trials. For purposes of analysis, a score was assigned to these *Ss* for the trials they did not complete which was the mean, for that trial, of the group to which they were assigned. The number of degrees of freedom was then decreased appropriately because of the resulting decrease in variability.

Following the procedure of the previous study (**7**), the acquisition expectancies were separated into thirds. The expectancy score for each *S* on each of the thirds consisted of the sum of his expectancy scores over that block of trials divided by the number of trials in the block. Therefore, the score for any third can be any value from 1 to 6, the same values as in the original rating scale. A score above 3 represents *S*'s expectancy that the machine will pay off on the next trial; a score below 3 represents *S*'s expectancy that the machine will not pay off. The overall course of expectancies as a function of percentage of reinforcement may be seen in Fig. 2. Figure 3 shows the overall course of expectancies as a function of the length of acquisition. An analysis of covariance was performed on the acquisition expectancies with the first third of the trials as the control variable and the last third as the experimental variable. The purpose of this analysis was to correct for any possible initial differences in expectancy. The interaction between percentage of reinforcement and the number of acquisition trials was found to be significant at the .01 level ($F = 3.38$ for 6 and 191 *df*). Using the interaction mean square as the error term, the effect of the percentage of reinforcement was significant at the .001 level ($F = 8.22$ for 2 and 191 *df*), but the effect of the number of acquisition trials was not significant. It would have been significant at the .05 level, however, if the within-group mean square had been used as the error term.

A trend analysis, with the first third and last third of the acquisition trials serving as points for the trend,

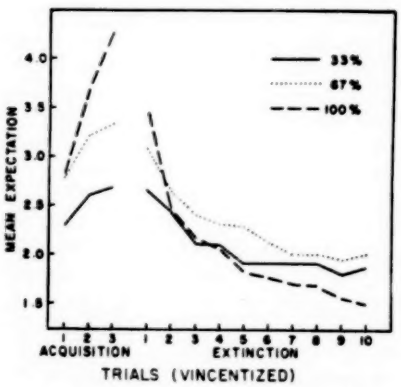

FIG. 2. The effect of percentage of reinforcement on the course of expectancies during acquisition and extinction.

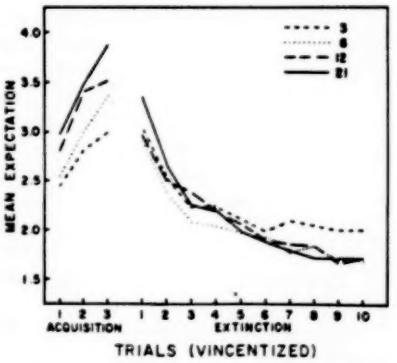

FIG. 3. The effect of the number of acquisition trials on the course of expectancies during acquisition and extinction.

was also performed with much the same results. There was a significant triple interaction among percentage, number of acquisition trials, and the points of the trend. The double interaction between percentage and the points of the trend remained significant even when the triple interaction was used as the error term, but the interaction between number of acquisition trials and the trend points was not significant.

A glance at Fig. 2 and 3 gives the meanings of these analyses. Figure 2 shows that 100% has a much greater effect on acquisition expectancies than 67%, which in turn has a greater effect than 33%. This is the meaning of the overall significance of percentage of reinforcement. In addition, it can be seen from Fig. 2 that percentage of reinforcement has a differential effect over trials, with the effect of 100% increasing more with a larger number of trials than the effect of either 67% or 33%. Figure 3 shows that the trend of increase during acquisition is not differential as a function of the number of acquisition trials, and this is the reason there is no significant interaction between number of trials and points of the trend. It is still possible, however, that the number of acquisition trials could have had an effect to which the analyses were not sensitive. The first third of the acquisition trials were grouped together for these analyses, and since there were, over all, a greatly different number of acquisition trials, the first third also represents a different number of trials. For the four groups there are 1, 2, 4, and 7 trials respectively in the first third of the trials. If the effect of the total number of trials occurred almost exclusively during these early trials—a very unlikely event—then the present analyses would be inadequate.

Extinction expectancies.—The expectancies for each *S* during extinction were separated into tenths and Fig. 2 and 3 show these Vicentized tenths through the course of extinction. An analysis of variance for trend was performed on these data with the last third of the acquisition trials, the second and the last tenth of the extinction trials serving as points in the trend. A summary of this analysis is presented in Table 1. Here it can be seen that percentage alone, and the interaction of percentage with the number of acquisition trials, were significant, while the number of acquisition trials alone was not significant. These findings are probably not too meaningful since it is already known that the groups differed significantly at the end of acquisition. What is meaningful, however, are the interactions of the various variables with trials, or trend points, since these interactions give the effect of differential changes in trend.

It can be seen that the Trials × Percentage interaction is significant at the .001 level, but that the trials by length interaction is not significant, nor is the triple interaction. This means that expectancies from the end of acquisition to the end of extinction change differentially only as a function of the percentage of reinforcement. Figure 3 shows the meaning of this differential change. Expectancies began the extinction series with the highest mean magnitude following 100% reinforcement, and they showed the greatest decrease for this same group. An analysis of variance performed on the mean of the last two blocks of extinction trials yielded a significant difference ($F = 3.54$ for 2 and 181

df) as a function of the percentage of reinforcement; no other significant differences were found. This significant F was due to the 100% group having lower expectancies than the other two.

Relation between the number of plays to extinction and expectancies.—In a previous report (7) it was indicated that the relationship between percentage of reinforcement and acquisition expectancies was essentially linear, and that the same linear relationship held between percentage of reinforcement and loss in expectation during extinction. One could therefore predict for that experiment that the curve relating percentage of reinforcement to mean log. plays to extinction would be very similar to curves relating last acquisition expectation or loss in expectancy to mean log. plays to extinction.

Figures 4 and 5 show the relation for the present experiment between percentage of reinforcement and the last acquisition expectations (Fig. 4) and between percentage of reinforcement and the overall loss in expectancy (Fig. 5). It can be seen that for the 21-trial group this relationship is quite linear, although not so much so for the overall loss in expectancy.

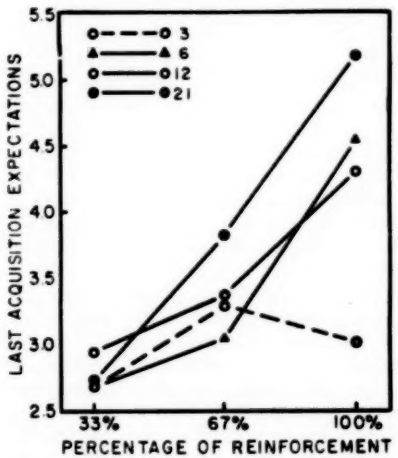

FIG. 4. The effect of percentage of reinforcement upon the last acquisition expectancies.

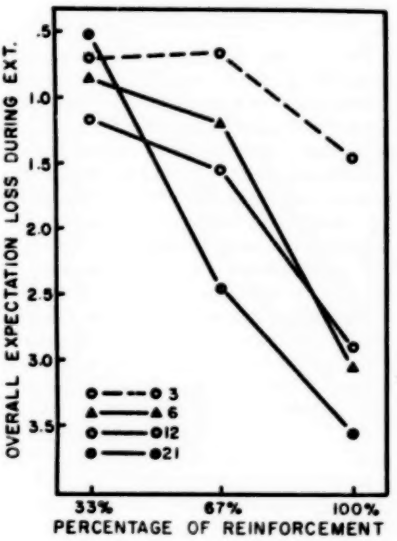

FIG. 5. The effect of percentage of reinforcement upon the overall loss of expectancies during extinction.

TABLE 1

ANALYSIS OF THE CHANGE OF EXPECTANCY FROM THE END OF ACQUISITION TO THE END OF EXTINCTION

Source	*df*	*MS*	*F*
Percentage (P)	2	10.89	5.70**
Acquisition trials (AT)	3	2.24	1.17
P × AT	6	4.46	2.34*
Between *Ss* (same group)	192	1.91	
Trials (T)	2	143.65	81.16***
T × P	4	19.56	11.05***
T × AT	6	3.45	1.95
T × AT × P	12	1.78	1.01
Pooled *Ss* × T	384	1.77	

* $P = .05$.
** $P = .01$.
*** $P = .001$.

Discussion

The major findings of the present experiment seem to be: (*a*) smaller percentages of reinforcement seem to result in a greater number of plays to extinction although the effect is not statistically reliable; (*b*) the larger the number of acquisition trials the fewer the number of trials to extinction; (*c*) expectancies increase differentially during acquisition as a direct function of the percentage of reinforcement so that at the end of acquisition, for the 21-trial acquisition group, expectancies approach a linear function of the percentage of reinforcement; (*d*) expectancies decrease differentially during extinction as a direct function of the percentage of reinforcement and this expectancy loss approaches linearity for the 21-trial acquisition group; and (*e*) there is no clear-cut statistical evidence that the number of acquisition trials has any effect upon expectancy.

A discrimination hypothesis has frequently been offered (**1, 3, 8**) as at least a partial explanation of the Humphreys' effect. The discrimination hypothesis maintains that extinction occurs when it is in some fashion clear to the organism that no more rewards are likely to occur. The discrimination hypothesis also maintains that the organism can respond to the acquisition series as a whole. Therefore, if the acquisition series is very similar to the extinction series the discrimination will be difficult and extinction will be prolonged. If the acquisition series is very dissimilar to the extinction series, discrimination will be easy and extinction will occur quickly. To make this hypothesis testable it is necessary to specify the conditions which make for easy and difficult discrimination.

The present experiment affords two variables of which ease of discrimination may be said to be a function. One variable is the percentage of reinforcement. It has been hypothesized that the smaller the percentage of reinforcement, the more similar will be the acquisition series to the extinction series, and the more difficult it will be to discriminate between the two. Thus more responses will occur during the extinction period. Such results have, of course, often been reported (**4**). Although the usual partial effect has not been found in the present study, the trend of Fig. 1A is highly similar to that of a similar experiment where it was found.

The other significant variable of this experiment of which discrimination could be said to be a function is the length of the acquisition series. One could reasonably predict that the longer the acquisition series is, the more the conditions of acquisition would stand out, up to a point, and the quicker would be the extinction. Thus there should be an inverse relation between length of acquisition series and plays to extinction. Present results confirm this prediction.

One would also predict that the group with 21 acquisition trials and 100% reinforcement would be the quickest to quit because this cell represents the extremes on an ease of discrimination dimension for both variables. The 33% and the 3-acquisition-trial group should play the longest because discrimination would be the most difficult. The data show that the 21 trial-100% group was the first to quit and the 3 trial-33% group was the next to last to quit.

These results seem to conflict with a considerable number of studies (**9, 12**) which show resistance to extinction to be a direct function of the number of acquisition trials. The resolution of the conflict seems to lie in the difference between learning and performance. Within limits, the accepted generalization that more trials lead to greater resistance to extinction holds for *learning*. Once a response has been greatly overlearned, e.g., the lever pulling response of this experiment, further trials serve simply to set the acquisition series off from the extinction series, and an inverse relationship between number of acquisition trials and resistance to extinction should result. The present data, then, are not inconsistent with the

two-process theory which has been discussed previously (**3, 7**), nor with some previous data (**5**).

We have here considered expectancies as a response measure, but it has been traditional (**11**) to consider expectancies as intervening variables that actually intervene between stimulus and response in the organism. Expectancies thus considered are believed to be determined by the stimulus variables and in turn to determine the responses. Because a previous experiment (**7**) indicated an approximately linear relationship between percentage of reinforcement and expectancies, and also an approximately linear relationship between expectancies and mean log. plays to extinction, it seemed reasonable to guess that these expectancies might be of the intervening type. The relationships of the present experiment probably are not sufficiently linear, or regular (see Fig. 4 and 5), to warrant the assumption that the expectancies are indices of an intervening variable. In addition, percentage of reinforcement had a significant effect upon the expectancies, both in acquisition and extinction, but it did not have a significant effect upon mean log. plays to extinction. This would seem to indicate that expectancies cannot be considered to intervene as a causal agent between percentage of reward and mean log. plays to extinction.

In an attempt to bring further evidence to bear on this point, an analysis of covariance was performed using the expectancies on the last acquisition trial as the control variable and mean log. plays to extinction as the experimental variable. The purpose of this analysis was to rule out the effects of expectancies while leaving any effect of percentage of reinforcement, not tied to expectancy, free to vary. The effect of the percentage of reinforcement remained nonsignificant, of course, and the analysis was not crucial on this main point. It did afford some interesting information, however, in that the within-groups and between-groups correlations of expectancy with log. plays to extinction could be easily determined. The correlation between the last acquisition trial expectancy and the mean log. plays to extinction was a − .58, significant at the .05 level for 10 *df*. Thus, as far as groups are concerned, the higher the last acquisition trial expectancy the lower the mean log. plays to extinction. That this is not the whole story, however, is shown by the within-group correlation which gives the relationship between expectancies and resistance to extinction for individuals. The within-groups correlation was only .027, which is far from significant. The relationship between expectancy and resistance to extinction found in this study, then, is a group effect and not an individual effect. A similar analysis was performed on previous and similar data (**7**) with the same result.

It has been pointed out several times before (**2, 10**) that group data and individual data do not always yield the same results. This study affords a graphic illustration of this point. What the status of expectancies can be, at least for the data presented here, is difficult to see. Certainly these expectancies cannot be conceived of as intra-individual phenomena.

Summary

In a factorial design, using an electric slot machine, 33%, 67%, and 100% reward during acquisition was combined with 3, 6, 12, and 21 acquisition plays. Following the last acquisition play, extinction began and no *S* was rewarded again although he was permitted to play as long as he desired. One response measure was the number of extinction plays, and a second was *S*'s stated expectancy of whether he expected to win or lose on the next play. The results indicated that:

1. The larger number of acquisition trials resulted in fewer plays to extinction.

2. The trend was for the smaller percentages of reinforcement to bring about a greater number of plays to extinction.

3. Expectancies increased differentially during acquisition as a direct function of the percentage of reinforcement.

4. Expectancies decreased differentially during extinction as a direct function of the percentage of reinforcement.

5. There was no clear-cut statistical evidence that the number of acquisition trials had any effect upon expectancies.

It was pointed out that these data are not inconsistent with a discrimination hypothesis; that the number of "acquisition" trials probably has a different effect upon a well learned response than upon a response just being learned; and that the relationship between expectancies and resistance to extinction is a group effect and not an individual effect.

References

1. Bitterman, M. E., Federson, W. E., & Tyler, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.
2. Estes, W. K. The problem of inference from curves based on group data. *Psychol. Bull.*, 1956, **53** 134–140
3. Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.
4. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
5. Lewis, D. J., & Duncan, C. P. The effect of partial reinforcement and length of acquisition-series upon resistance to extinction of a motor and a verbal response. *Amer. J. Psychol.*, 1956, **69**, 644–646.
6. Lewis, D. J., & Duncan, C. P. Effect of different percentages of money reward on extinction of a lever-pulling response. *J. exp. Psychol*, 1956, **52**, 23–27.
7. Lewis, D. J., & Duncan, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and amount of reward. *J. exp. Psychol*, 1957, **54**, 115–120.
8. Mowrer, O. H., & Jones, H. J. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **35**, 293–311.
9. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.
10. Spence, K. W. *Behavior theory and conditioning.* Yale Univer. Press., 1956.
11. Tolman, E. C. Purposive behavior in animals and men. New York: Appleton-Century, 1932.
12. Williams, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–521.

(Received January 25, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

EFFECTS OF MEDIATED ASSOCIATIONS ON TRANSFER IN PAIRED-ASSOCIATE LEARNING

KATHRYN J. NORCROSS AND CHARLES C. SPIKER

Iowa Child Welfare Research Station

Several experimenters have demonstrated that a given response (R_o) may be elicited by a stimulus (S_1) to which R_o has never been conditioned, even when S_1 is physically dissimilar to any stimulus to which R_o has been conditioned (**1, 3, 6, 8, 11**). These results have been explained in terms of the principle of secondary, or mediated, generalization (**4**). Briefly, the principle states that the conditioning of a response, R_a, to each of several dissimilar stimuli (S_1, S_2, etc.) assures that on subsequent presentations of these stimuli, the stimulation, s_a, produced by R_a, will be present. Any other response, R_o, subsequently conditioned to one of these stimuli (e.g., S_2) will also be conditioned to s_a. Thus, if any other of the stimuli (S_1, etc.) is presented, R_a will tend to occur, together with s_a, and s_a will tend to elicit R_o.

The principle of mediated generalization also appears applicable to the phenomenon of mediated associations in paired-associate verbal learning. The paradigm for mediated associations provides for two preliminary lists and a third test list administered in an A-B, B-C, A-C sequence. A schematization of the analysis of this situation is presented in Fig. 1. In List I, S is presented with visual stimulus A, to which he makes the (overt or covert) verbalization R_A with attendant stimulation, s_a. S_B is presented in close spatial and temporal relationship with S_A and elicits R_B, which in turn produces s_b. The experimental conditioning that occurs is represented by the dotted lines in Fig. 1. As a result of List II experience, s_b comes to evoke R_C. The subsequent learning of R_C in response to S_A is facilitated through mediation by s_b.

An early series of experiments on mediated association with some positive findings was reported by Peters (**9**). Associative facilitation was investigated by Irwin (**5**) with a B-C, A-B, A-C sequence for an experimental group as compared to a D-C, A-B, A-C control sequence. A significant difference between the two groups in the learning of the A-C list was found, with the group given the B-C list superior. Bugelski and Scharlock (**2**) used an experimental design in which the third A-C list was constructed from previously associated nonsense syllables in such a way that experimental pairs were made up of an A syllable and a C syllable that had been associated with a common B

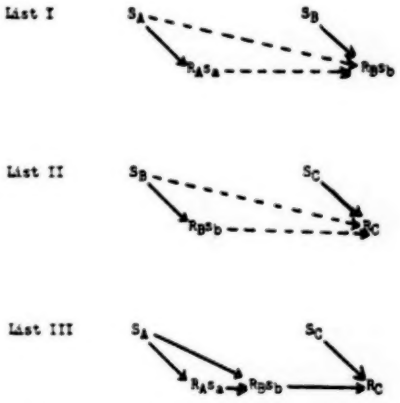

FIG. 1. A schematic analysis of mediated association in paired-associate learning.

syllable; control pairs were composed of the remaining A and C syllables paired at random so that none had a common specific B associate. A significant difference was obtained in the mean rank order of learning the two types of pairs. Russell and Storms (**10**) have recently demonstrated mediational effects with experimentally established learning and inferred pre-existing language habits.

The differences that obtained in the latter experiments (**2, 10**) may have resulted from a combination of positive transfer for the experimental condition and negative transfer in the control condition, since mediated but incorrect response tendencies presumably exist for the "control" items. The present experiments attempt to demonstrate both positive and negative transfer in learning the third list in the mediated association design.

Experiment I

Method

Apparatus.—The apparatus used is an electrically operated visual stimulus presentation device.[1] It consists of a gray metal box 14 × 14 × 14 in. The front face of the apparatus contains a 3 × 6 in. aperture. A mechanical arrangement delivers to the aperture a 3½ × 6 in. plastic card from the bottom of a stack of cards. Mounted on each card was a pair of pictures taken from the *Golden Play Book of Picture Stamps for the Very Young.* Each picture was a colored representation of a single familiar object; e.g., a dog, horse, ice cream cone, etc. Black cloth curtains on the aperture serve as shutters to expose or conceal independently the left and right pictures on the stimulus cards. Following a presentation, the curtains close simultaneously and the card is released from the aperture and returned to the top of the stack by a conveyor belt. A system of electronic timers controls the rate of presenting the cards and the exposure times of the stimulus and response pictures.

[1] The apparatus was especially constructed by the Hunter Mfg. Co., Iowa City, Iowa. The authors are indebted to Dr. Alfred Castaneda for aid in formulating the specifications for the apparatus.

TABLE 1

Schematic Illustration of the Design for Exp. I

List I		List II		List III	
S	*R*	*S*	*R*	*S*	*R*
1	7	7	13	1	13*
2	8	8	14	2	14*
3	9	9	15	3	16**
4	10	10	16	4	15**
5	11	19	17	5	17†
6	12	20	18	6	18†

* Pairs in List III for which facilitation is predicted.
** Pairs in List III for which interference is predicted.
† Control pairs in List III.

Experimental design.—Each *S* learned three lists of stimulus-response pairs. Each list consisted of six pairs of pictures. In order that *S* might more easily comprehend the task, List I was constructed of picture pairs which were highly associated; Lists II and III were not so constructed. Table 1 illustrates the design, which provides that each *S* serve under each of three experimental conditions. For two pairs in List III, the mediational chain is expected to facilitate learning. For two other pairs, the mediational chain is expected to elicit incorrect responses and thus interfere with learning. Two other pairs have no experimentally established mediating tendencies, since the response items paired with the stimuli in List I were not presented in List II.

Subjects.—The *S*s were 30 kindergarten children.[2] Ten *S*s were randomly assigned to each of three groups. These three groups were designed to control for possible differential difficulty of the stimulus-response pairs in List III. All groups received the same form of Lists I and III. The groups differed with respect to which of three forms of List II they received. These three forms were constructed so that for a given List III pair, facilitation was predicted in one group, interference in another, and neither mediated facilitation nor interference for the third group.

Procedure.—Each *S* was given three experimental sessions separated by from two to seven

[2] The authors are indebted to Hale C. Reid, Director of Curriculum and Instructions, for permission to use subjects in the Cedar Rapids Public Schools. Special thanks are also due to Ray Churchill, Principal, and to the staff of Harrison School in Cedar Rapids, for administrative assistance and cooperation in the collection of the data of Exp. I.

days. After the child had been brought by *E* from the classroom to the experimental room in order to play a "guessing game with pictures," he was seated facing the aperture in which the stimulus cards were presented. The *E* was seated immediately beside *S* in order to control the apparatus, give appropriate verbal instructions to establish understanding of the procedure, and to record responses made in the anticipation interval.

At the beginning of training, the right half of the aperture was opened and the six response members of List I were shown once each. The *S* was asked to name them. Any reasonable name given was accepted; if necessary, a name was supplied. The same procedure was followed as the stimulus members were presented in the left half of the aperture. A second naming trial was then given with the response members. During the first paired presentation, the stimulus side of the pair was exposed while *E* asked *S* each time what he thought was behind the curtain on the other side. The stimulus and response items were then simultaneously exposed while *S* gave the response name, either voluntarily or upon request. After the first trial, *E* instructed *S* to try to guess which picture was behind the curtain each time. Presentation was continued until a learning criterion of two consecutive correct trials had been reached.

Three varying orders of presentation were used for all lists in order to control for serial learning. A 6-sec. anticipation period was followed by a 5-sec. period of joint presentation of the stimulus and response items. The between pair interval was 2.2 sec. All verbal responses occurring in the anticipation interval were recorded verbatim throughout the experiment.

On the second experimental day, List II was administered in the same way as was List I, except that a simultaneous presentation of both the stimulus and response members was given rather than the second presentation of the response items alone.

On the third day, Lists I and II were each reviewed to criterion. A short rest period of approximately 2 min. followed. List III was then introduced in the same way as List I on the first day. Five trials on List III were given, terminating the experiment.

Results and Discussion

The mean number of trials to criterion for all Ss on List I was 2.13. The means of the number of trials for the various forms of List II did not differ significantly; the overall mean was 3.47. The more rapid learning of List I than of List II is probably due to the higher associations within pairs for List I.

An analysis of variance of correct anticipations on List III is summarized in Table 2. In this analysis, the main effect of control groups appears as a between-*S* factor and the main effect of conditions as an intra-*S* factor (**7,** p. 267, Type I). This analysis revealed that the means for the groups designed as controls for possible differential difficulty of the picture pairs did not differ significantly. The means for the facilitation, control, and interference conditions differed at better than the .001 level of significance. For all 30 *S*s combined, the mean numbers of correct responses in five trials were 7.40 for the facilitation, 4.83 for the control, and 4.40 for the interference conditions. Individual *t* tests showed that the mean differences for the facilitation and control comparison and for the facilitation and interference comparison were significant at better than the .001 level. The difference between control and interference means, although in the predicted direction, was not significant.

According to the mediation hypothesis, the learning of Lists I and

TABLE 2

ANALYSIS OF VARIANCE OF CORRECT RESPONSES ON LIST III (EXP. I)

Source	*df*	*MS*	*F*
Between *S*s	29		
Control groups (G)	2	2.98	<1.00
Error (b)	27	8.48	
Within *S*s	60		
Conditions (C)	2	78.88	28.89*
G × C	4	2.48	<1.00
Error (w)	54	2.73	
Total	89		

* $P < .001$.

II should have set up a tendency for the stimulus member of each interference pair in List III to elicit the response that was correct for the other pair. Thus, the hypothesized interference effects may be further tested by comparing the number of these within-set errors for the two interference pairs with the number that occurred for the two control pairs, the latter reflecting the chance frequency of such reversals. The mean number of reversals per *S* for the interference condition was 1.50 and for the control condition, .20. A related *t* test indicated that the mean difference was significant at the .001 level.

Thus, the results of Exp. I indicate clear-cut positive transfer with a mediating facilitation sequence. Significant overall negative transfer was not shown, but the interference condition produced significantly more within-set errors than did the control condition. This latter finding is consistent with the hypothesis that specific incorrect responses are mediated under the interference condition.

Experiment II

Experiment II was conducted to investigate further the interference effects due to the mediational sequence. It was thought that by increasing the number of pairs in the interference and control conditions and by omitting the facilitation condition, a significant overall negative transfer might be obtained for the interference condition. The apparatus was the same as that used in Exp. I. The stimulus materials were of the same type as those in Exp. I.

Method

Experimental design.—Each *S* learned three lists of stimulus-response pairs with six pairs of pictures per list, List I again being constructed of high association pairs. Table 3 illustrates the design. List III was composed of three control pairs and three pairs for which mediated interference was predicted. Thus, each *S* served in both the interference and control conditions.

TABLE 3

Schematic Illustration of the Design for Exp. II

List I		List II		List III	
S	*R*	*S*	*R*	*S*	*R*
1	7	7	13	1	14*
2	8	8	14	2	15*
3	9	9	15	3	13*
4	10	19	16	4	16**
5	11	20	17	5	17**
6	12	21	18	6	18**

* Pairs for which interference is predicted in List III.
** Control pairs in List III.

Subjects.—The *S*s were 24 first grade children.[3] Twelve *S*s were randomly assigned to each of two groups. These groups were designed to control for possible differential difficulty of the stimulus-response pairs in List III. Both groups were given identical forms of Lists I and III; the groups differed only with respect to which of two forms of List II they received. These two forms were constructed so that any given pair in List III would be a control pair for one group and an interference pair for the other.

Procedure.—The procedure for Exp. II was essentially the same as that for Exp. I. In order to adjust the difficulty of the tasks to the older first-grade *S*s, the anticipation interval was reduced from 6 sec. to 4 sec., and the joint presentation interval was reduced from 5 sec. to 2 sec.

Results and Discussion

The mean number of trials to criterion for all *S*s on List I was 1.67. The means of the number of trials for the two forms of List II did not differ significantly; the combined mean was 2.38. As in Exp. I, it should be noted that the pairs in List I were of highly associated items.

[3] Appreciation is expressed to Dr. Jerry Kuhn, Principal, University Elementary School, Iowa City, Iowa, for his permission to use the first grade class and for providing experimental facilities for Exp. II.

TABLE 4

ANALYSIS OF VARIANCE OF CORRECT RESPONSES ON LIST III (EXP. II)

Source	*df*	*MS*	*F*
Between *Ss*	23		
Control groups (G)	1	35.02	2.33
Error (b)	22	15.07	
Within *Ss*	24		
Conditions (C)	1	31.68	7.90*
G × C	1	2.53	<1.00
Error (w)	22	4.01	
Total	47		

* $P = .01$.

Table 4 presents a summary of the analysis of variance of correct anticipations in five trials on List III. The stimulus control groups did not differ significantly. The mean number of correct anticipations for all 24 *Ss* was 7.83 for the control condition and 6.21 for the interference condition. The difference is significant at about the 1% level.

As in Exp. I, the within-set errors were tabulated for the interference and control pairs. There were 37 such errors for the interference condition and only 9 for the control. The mean difference of 1.17 yields a t of 3.19 ($P < .01$).

Experiment II thus provides statistically significant evidence that the conditions designed to produce the mediation of incorrect responses result in overall negative transfer in learning List III. When both experiments are considered, it appears that, under appropriate conditions, both positive and negative associative transfer is a consequence of mediated associations.

It is of interest to compare the results of the present experiments with those of Bugelski and Scharlock (**2**). Their experiment is like these in that the mediational tendencies were experimentally manipulated. Their experimental condition corresponds to the facilitation condition of Exp. I, while their control condition is analogous to the interference conditions of Exp. I and II. They found no significant differences in the mean or median number of trials to criterion for the two conditions; a significant difference was found in the mean rank order of learning the pairs in the two conditions. In the present investigation, the use of a control condition, in which no mediational tendencies existed that did not also exist for the other conditions, permitted the separation of the facilitation and interference effects.

SUMMARY

This investigation proposed to study the effects of mediated associations on the positive and negative associative transfer in the paired-associate learning of kindergarten and first-grade *Ss*.

In Exp. I, a different list of six stimulus-response picture pairs was learned by each of 30 *Ss* on each of three days. The learning sequence was such that mediation would be expected to have facilitating, interfering, and no effects for each *S* in the learning of different pairs of List III. A statistically significant difference was found between the mean number of correct anticipations for the control and facilitation conditions during List III learning. The difference between the means for the control and interference conditions, although in the predicted direction, was not statistically significant.

In Exp. II, each of 24 *Ss* learned a different list of six stimulus-response picture pairs on each of three days. The first two lists were arranged so that, in the learning of List III, interference effects were predicted for three pairs relative to the other three pairs. A statistically significant difference was found in the mean number of correct anticipations for the interference and control conditions.

It was concluded that, under the appropriate conditions, mediated associations can produce both positive and negative associative transfer.

REFERENCES

1. BIRGE, J. The role of verbal responses in transfer. Unpublished doctor's dissertation, Yale Univer., 1941.
2. BUGELSKI, B. R., & SCHARLOCK, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.
3. EISMAN, B. S. Attitude formation: the development of a color-preference response through mediated generalization.

J. abnorm. soc. Psychol., 1955, **50,** 321–326.

4. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46,** 9–30.
5. Irwin, Irl A. Associated facilitation as a function of the strength of the associative connection between first and second list response words. Unpublished doctor's dissertation, State Univer. Iowa, 1951.
6. Jeffrey, W. E. The effects of verbal and non-verbal responses in mediating an instrumental act. *J. exp. Psychol.*, 1953, **45,** 327–333.
7. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
8. Murdock, B. B., Jr. Effects of failure and retroactive inhibition in mediating an instrumental act. *J. exp. Psychol.*, 1952, **44,** 156–164.
9. Peters, H. N. Mediate association. *J. exp. Psychol.*, 1935, **18,** 20–48.
10. Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49,** 287–293.
11. Shepard, W. O. Mediated generalization with high interstimulus similarity. Unpublished M.A. thesis, State Univer. Iowa, 1953.
12. Shipley, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12,** 337–357.

(Received February 11, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

TRANSFER EFFECTS FROM A SINGLE TO A DOUBLE INTEGRAL TRACKING SYSTEM[1]

GEORGE E. BRIGGS, PAUL M. FITTS
The Ohio State University
AND HARRY P. BAHRICK
Ohio Wesleyan University

The empirical study of complex skill learning and the generation of training principles require a dimensional analysis of task variables. Such an analysis should include identification of those task characteristics which define (*a*) the perceptual processes, (*b*) the nature of the response sequences, and (*c*) the amount and form of feedback information available to or required by *S* in his operation of a skill system. This study is the third of a series (**2**, **3**) which offers information on this type of analysis for a relatively complex, two-dimensional tracking system.

Specifically, the present study was concerned with the lag characteristics of a physical system, in this case the number of integral transformations which the machine imposed on signals from *S*'s control device. As the number of such transformations increases, the pattern of response and the perceptual role demanded of *S* become increasingly complex. To illustrate, the patterns of response shown in Fig. 1 represent one set of a family of such patterns required to compensate perfectly for a step-function input when control device signals are subjected to zero, one, two, and three stages of integration. The time and amplitude axes were purposely labelled dimensionless because the time scale of such idealized patterns as well as the time scale of the machine output will vary as a function of the constants of integration (system sensitivity) and as a function of the amplitude and duration of *S*'s control movements. These latter characteristics in turn are functions of both the control limits of displacement and the experience of *S* with a particular system.

The relation among the four response patterns shown in Fig. 1 is quite straightforward and indicates the nature of the roles required of *S* in his operation of these systems: each response pattern is the first derivative of the preceding pattern. Therefore, the perceptual task confronting *S* may be said to be analogous to the operations performed by a simple amplifier, a single, dual, or triple stage differentiator, depending on whether control is exercised through zero, one, two, or three stages of integration. As shown in Fig. 1, this order of control may be identified as positional, velocity (or rate), acceleration, and Δ acceleration.

It is apparent that the order of the control system provides a convenient index to task complexity, and considerable research effort has been

[1] This research was conducted in the Laboratory of Aviation Psychology of The Ohio State University and was supported in part by the United States Air Force under Contract No. AF 18(600)-1201 with the OSU Research Foundation, monitored by Dr. J. A. Adams, who served as Scientific Officer for the AF Personnel and Training Research Center. Permission is granted for reproduction, publication, use, and disposal in whole or in part by or for the United States Government.

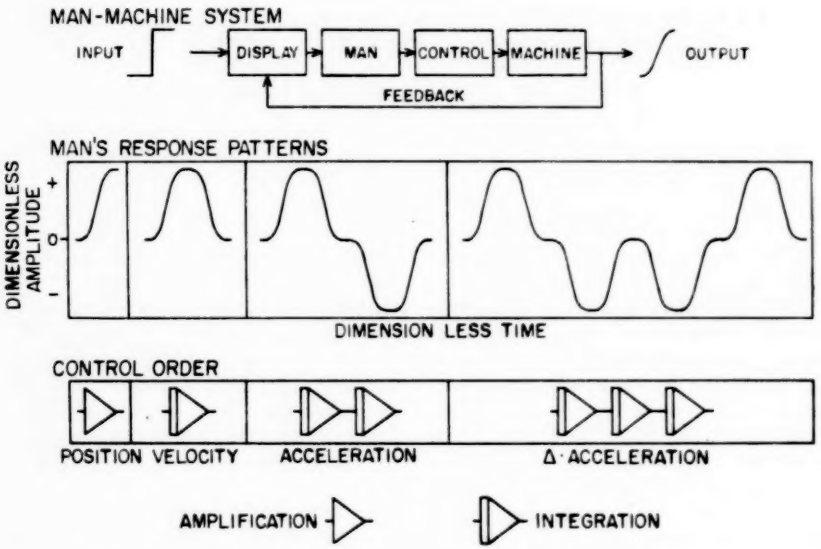

FIG. 1. Idealized response patterns to a step input as a function of the number of integral transformations performed on control signals. After Hick and Bates (6).

directed toward specifying the capacity of *S* at terminal levels of training as a function of control order. Concurrently, research has attempted to discover display and control design principles which will permit higher levels of performance earlier in the training period (1, 6). Birmingham and Taylor (1), for example, propose methods of "quickening" a tracking display such that *S* must still execute complex patterns of response, but in so doing he is not required to estimate higher derivatives of the stimulus. Further, research on velocity (8) and on acceleration (7) tracking systems has demonstrated relatively high degrees of positive transfer with non-quickened displays following extended practice on quickened displays.

Relatively little is known regarding transfer of training from one order of control to another. The only relevant data on this point come from Lincoln (8) who measured transfer performance on a velocity system following extended training with a positional control tracking system. Transfer effect was positive but at a relatively low level.

The present study was intended to supply additional information on transfer from one order of control to another. Transfer effects were measured on an acceleration system following several amounts of training on a velocity control system. The tracking task was two dimensional, but the number of lag terms in only one dimension was varied. A non-quickened, compensatory tracking display was employed with a step-function input.

METHOD

Apparatus.—The OSU Pilot Training Research Simulator (5) served as the skill task.

This apparatus utilizes an analog computer which simulates the general tracking conditions required by an interceptor aircraft with a radar fire-control system. The *S* views a 5 in. cathode ray tube (CRT) display which provides target and aircraft attitude information for compensatory tracking. Before the start of a tracking trial the target signal is given an initial displacement from the center of the CRT, an area defined by a circular reticle, to one of the four quadrants of the CRT. The system input, therefore, is a step function.

The *S* is instructed to return the target signal to the reticle and to maintain that position throughout the remainder of the trial. This is accomplished by the use of a two-dimensional, spring centered control column which generates two voltages: lateral displacements produce a voltage, *X*, which is subjected to two stages of integration, the first representing rate of roll, the second simulating rate of change in heading proportional to the assumed bank angle; fore and aft deflections of the control generate a voltage, *Y*, which is transformed by an exponential oscillatory function plus one stage of integration, the combination simulating an underdamped response of the system to pitch corrections. A flow diagram of these transformations is shown in Fig. 2. The vertical position and angle of incidence of an artificial horizon on the display are determined by the outputs of the first transformations in the *X* and in the *Y* components of the system. The operation shown as 1/T is employed to simulate an increase in system sensitivity as the interceptor closes on the target. The value of *T* is continuously variable from 40 at the start of a trial, to 1 at the end.

Figure 2 defines the system transformations present in the "whole" or transfer task. The "part" or training task consisted of the same system elements with the exception that the first stage of integration in the *X* channel was replaced by an amplifier of comparable gain. This produced a simplified velocity system for azimuth corrections during the training trials while the transfer task may be classed as an acceleration system for reduction of *X* error. As indicated in Fig. 1, the pattern of lateral responses demanded of *S* during training represented a considerable simplification over the response movements required during transfer. The pattern of response for reduction of *Y* error remained unchanged from training to transfer tasks. Since there was no interaction between the two system components (no loss of lift in bank corrections), it was possible for *S* to make azimuth and pitch corrections independently.

The major characteristics of the CRT display

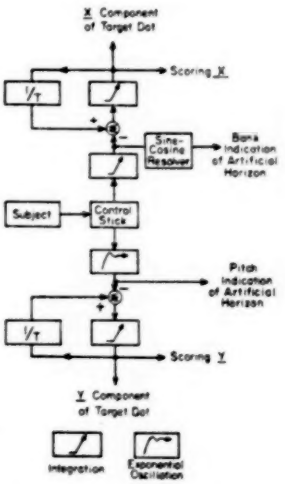

FIG. 2. A flow diagram of major transformations and operations performed on control signals in the transfer (acceleration) tracking system. See text for description of part-task training system.

were essentially the same during training and transfer sessions. In particular, the relation of artificial horizon degree and direction of inclination to direction and rate of target movement remained unchanged. Further, the direction of motion relations of lateral control column displacements and artificial horizon inclination also remained constant. In fact, the only displayed difference in training and transfer tasks was in the rate of rotation of artificial horizon following control column displacement.

Tracking performance was measured in rectangular coordinates. Analog squaring and integrating circuits provided indices of integrated error squared for *X* and for *Y* dimensions, error being defined as the momentary difference, in volts, between target position and the center of the circular reticle. These scores were taken over the final 20 sec. of each 40-sec. trial, and division of the integrated error square metric by 20 provided the basic data of the study: mean square error (MS_x and MS_y).

Subjects and procedures.—Fifty-six freshmen and sophomore students, volunteers from AFROTC classes at The Ohio State University, served in this experiment. Each *S* participated in five daily sessions of approximately 45 min. and was paid hourly wages. There were 10 tracking trials during the first session and 20 trials per session thereafter. Trial duration was

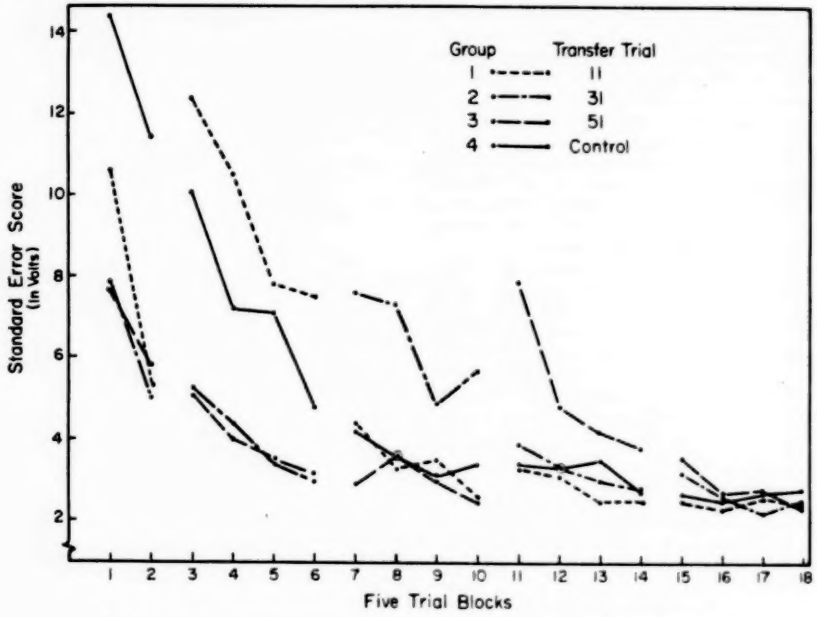

FIG. 3. Training and transfer performance levels as indicated by combined mean square error measures: $(MS_x + MS_y)^{\frac{1}{2}}$.

40 sec. with 20-sec. rest between trials and 5-min. rest between blocks of five trials. The *S*s served in pairs alternating tracking and observing in five-trial blocks.

Pairs of *S*s were assigned in a non-systematic manner to one control and three experimental groups where the amount of practice on the velocity system defined groups. Groups 1, 2, and 3 received 10, 30, and 50 trials, respectively, on the single integral task prior to transfer to the acceleration system. Group 4 received practice only on the acceleration system. There was a total of 90 tracking trials for each *S*. The data for three *S*s were invalidated by temporary equipment malfunction; the *N*'s for Groups 1 through 4 were 14, 12, 14, and 13, respectively.

RESULTS

Figure 3 presents the results of the training and transfer sessions while Table 1 lists the data of the transfer trials for each group. These

TABLE 1

AVERAGE STANDARD ERROR SCORES, IN VOLTS, FOR CONTROL AND EXPERIMENTAL GROUPS AT THE TIMES OF TRANSFER

Group	Transfer Trial Blocks											
	3	4	5	6	7	8	9	10	11	12	13	14
1	12.41	10.53	7.71	7.53								
2					7.58	7.28	4.88	5.69				
3									7.86	4.85	4.17	3.82
4	10.41	7.21	7.11	4.84	4.16	3.64	3.08	3.38	3.45	3.28	3.52	2.71

data are based on a combination of the two mean square error scores for each S on each trial: $(MS_x + MS_y)^{\frac{1}{2}}$. The average of these error measures was taken over five-trial blocks.

The superiority of the three experimental groups over the control group in Session 1 confirms the expectation that a velocity system would permit a higher initial level of performance than would an acceleration system. However, the inferiority of tracking through two integrations (for reduction of X error) was overcome by practice as Group 4 did attain a proficiency level during Session 3 comparable to that of Group 3.

The transfer performance of Groups 1, 2, and 3 on Sessions 2, 3, and 4, respectively, may be evaluated in several ways. First, as compared to the initial performance level of the control group in Session 1, the experimental groups show increasing positive transfer as a function of the amount of part-task training. A convenient index to express this transfer is $\frac{C_i - E_t}{C_i - TL} \times 100$, where C_i is the average performance level of Group 4 on Blocks 1 and 2; E_t is the average transfer performance levels of Groups 1, 2, and 3 on Blocks 3 and 4, 7 and 8, and 11 and 12, respectively; and TL is the average terminal performance level of all groups over Blocks 17 and 18. The value calculated for C_i is 12.88 v., while TL is 2.52 v. Values of E_t are 11.47, 7.43, and 6.36 v. for Groups 1, 2, and 3, respectively. This particular transfer index gives the relative amount of whole-task learning achieved by part-task practice, and therefore these are analogous to savings scores. The indices for Groups 1, 2, and 3 are 14%, 53%, and 63%, respectively.

Another comparison of interest is the number of trials required to attain a given proficiency level on the transfer task. The level of 4.3 v. suggests itself: this represents two-thirds the difference between C_i and TL. Figure 3 shows that Groups 4 and 1 achieved this criterion after 30 and 20 transfer trials, respectively (plus the 24-hr. interval between Sessions 2 and 3), while Group 2 reached the level after 20 transfer trials (plus the 24-hr. interval between Sessions 3 and 4), and Group 3 required approximately 15 transfer trials. These and the above savings scores suggest, then, that not only does practice on a more simple velocity system increase the initial tracking level following transfer to a more complex acceleration system, but also attainment of a specified level of performance on the whole task is facilitated as the amount of part-task training, on this particular task dimension (lag), is increased. Obviously, there is a limit to the amount of such savings, and the results of an analysis of variance (see below) suggest that Group 3 has closely approached such a limit.

Finally, one may evaluate transfer effects by relating the levels of the experimental groups to the control group at the times of transfer. A convenient index for this comparison is $\frac{C_i - E_t}{C_i - C_t} \times 100$, where C_i and E_t are as defined above, and C_t is the average levels of Group 4 at Blocks 3 and 4, at Blocks 7 and 8, and at Blocks 11 and 12 (8.81, 3.90, and 3.36 v., respectively). When so expressed, Group 1 shows 35% transfer while Groups 2 and 3 show 61% and 68%, respectively. Again, increasing positive transfer resulted from an increase in amount of part-task practice.

Table 2 lists the results of an analysis of variance for repeated

TABLE 2

ANALYSIS OF VARIANCE FOR THE DATA OF SESSIONS 2, 3, AND 4

Source	*df*	*MS*	*F*
Groups (G)	3	45.79	—
Ss within Groups (Ss/G)	49	69.76	
Sessions (D)	2	426.51	17.28*
D × G	6	266.12	10.78*
D × Ss/G	98	24.68	
Blocks (B)	3	144.11	46.04*
B × G	9	2.34	—
B × Ss/G	147	3.13	
B × D	6	16.45	5.24*
B × D × G	18	10.61	3.38*
B × D × Ss/G	294	3.14	

* $P < .05$.

measures for the data of Sessions 2, 3, and 4. The significant interaction of Blocks × Sessions × Groups is of particular interest as it justifies individual comparisons between control and experimental groups at the times of transfer. The results of these latter comparisons revealed significant differences ($P < .05$) between Group 4 and Groups 1, 2, and 3 on Blocks 3 and 4, 7 and 8, and 11 and 12, respectively.

Further, a comparison of group means for Sessions 2, 3, and 4 is justified by the significance of the interaction Sessions × Groups. The results of individual comparisons revealed statistically significant ($P < .05$) differences between Group 1 on Session 2 and Groups 2 and 3 on Sessions 3 and 4, respectively. However, the difference between Groups 2 and 3 on Sessions 3 and 4, respectively, did not approach significance at the .05 level. Thus, it appears that Group 3 has approached maximum transfer in this particular task, and additional velocity system training beyond 50 trials would not be expected to add materially to the values of the transfer indices noted here for Group 3.

DISCUSSION

The trend of the transfer indices noted above is in relatively close agreement with that reported by Gagne and Foster from an experiment of similar design employing discrimination learning procedures (**4**). In that study, training consisted of 0, 10, 30, and 50 trials of switch closures to color or to position of stimulus lights, while transfer involved similar switch closure responses differentially to both position and color of stimuli. Increasing positive transfer was obtained as a function of amount of part-task training.

It is to be noted that both the stimuli and the responses utilized by Gagne and Foster during transfer were quite similar to, but more complex than, their counterparts during training. Much the same description applies to the present study: as indicated above, the tracking display was essentially the same in appearance during training and transfer sessions, and Fig. 1 indicates that while the response pattern required in using an acceleration system is more complex than that required with a velocity control, the two are quite similar in their essential forms. Indeed, the expected acceleration-control response pattern can be described as two velocity-control patterns in serial order with a change in polarity as the major difference between the two segments. With stimuli and responses during transfer similar to their counterparts in training, S-R theory (**9**, p. 530 ff.) predicts, of course, some degree of positive transfer effects. As such, the results of both the Gagne and Foster and the present studies were not unexpected.

Therefore, on a gross descriptive level at least, the Gagne and Foster task and the present experimental situation were quite similar with respect to the change in stimuli and responses from training to transfer. This suggests an interesting parallel between transfer effects when the transfer task requires,

primarily, a change in the perceptual aspects, as in Gagne and Foster, and when transfer involves a change both in the perceptual aspects and in the patterning and precision of response, as in the present study. The responses required in that earlier investigation merely involved deflection of the appropriate one of four switches, and as such *S* was not required to respond with a precise pattern of movements on either training or transfer tasks. This is in contrast to the present tracking tasks which placed considerable demands on a precise pattern of control deflections, especially during transfer.

Apparently, this requirement of precise patterning of response does not of itself introduce a major change in the trend of the transfer indices as a function of amount of part-task practice. Thus, it is possible to utilize theory developed on the basis of discrete S-R transfer of training problems for prediction with certain of the task variables associated with and seemingly unique to the complex and continuous motor skill situations. As shown here, when the effects of additional lag terms in a tracking system are abstracted to stimulus-response terminology, S-R theory enjoys sufficient generality to predict the sign (positive) and the trend of transfer effects.

Supplementary to S-R theory, the computer network and servo theories, as for example Birmingham and Taylor (**1**), make it possible to add to these predictions of transfer effects in motor skill. To illustrate, it was not expected that the transfer indices, noted above, would approach or exceed the 100% level. In a previous report (**2**) the authors proposed that for maximum proficiency during transfer, the training task should contain the major component interactions and transformations of the transfer system. Changes either in the number of transformations, as done here, or in the number or nature of system interactions from training to transfer tasks affect a change both in the perceptual role required of *S* and in the pattern and precision of response. Thus it appears that 68% transfer, as found here for Group 3, is quite high.

While the predictive power of existing theory is somewhat less than desired, it is to be expected that the methodology and research suggested by servo theory will add significantly to our understanding of and theory for complex skill learning.

Summary

This research was concerned with transfer effects from a part-task, which involved tracking through a single stage of integration, in the *X* axis of a two-dimensional task, to a system requiring tracking through a double integration. Amount of practice on the single integral system was varied for 0, 10, 30, and 50 trials. The basic task was a two-dimensional compensatory tracking situation with a step-function input.

Performance on the transfer task was evaluated in several ways. When the experimental groups at transfer are compared to the initial level of the control group, there was increasing positive transfer as a function of number of trials on the training task. In terms of standard error scores there was 14%, 53% and 63% transfer for the 10, 30, and 50 trial groups. Examination of the performance functions during transfer showed that not only did part-task practice increase the initial transfer performance level, as indicated by the above transfer scores, but also attainment of criterial levels on the whole task was facilitated as amount of part-task training was increased. Finally, when transfer comparisons were made between experimental and control groups at the time of transfer, the transfer indices for the three experimental groups were 35%, 61%, and 68%.

References

1. Birmingham, H. P., & Taylor, F. V. A human engineering approach to the design of man-operated continuous control systems. Washington, D. C.: Naval Res. Lab. Rep. No. 4333, April 1954.
2. Briggs, G. E., Fitts, P. M., & Bahrick, H. P. Learning and performance in a complex tracking task as a function of visual noise. *J. exp. Psychol.*, 1957, **53**, 379–387.
3. Briggs, G. E., Fitts, P. M., & Bahrick, H. P. Effects of force and amplitude cues on learning and performance in a complex tracking task. *J. exp. Psychol.*, 1957, **54**, 262–268.
4. Gagne, R. M., & Foster, H. Transfer of training from practice on components

in a motor skill. *J. exp. Psychol.*, 1949, **39**, 47–68.

5. Harter, G. W., & Fitts, P. M. The functional simulation of complex systems by means of an analog computer, with the F-86D, D-4 system as a specific example. AFPTRC, Res. Rep., 56–133, Part 1, Dec. 1956.
6. Hick, W. E., & Bates, J. A. V. The human operator of control mechanisms. London: Ministry of Supply, Permanent Records of Research and Development, No. 17–204, May 1950.
7. Holland, J. G., & Henson, J. B. Transfer of training between quickened and unquickened tracking systems. *J. appl. Psychol.*, 1956, **40**, 362–366.
8. Lincoln, R. S. Visual tracking: III. The instrumental dimension of motion in relation to tracking accuracy. *J. appl. Psychol.*, 1953, **37**, 489–493.
9. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

(Received February 14, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

SOME INFORMATIONAL ASPECTS OF FORM DISCRIMINATION

GILBERT K. KRULEE

Case Institute of Technology[1]

In a previously reported series of experiments (**2, 3**), two general hypotheses concerning the determination of thresholds for multiple-unit displays were examined. In this report, their implications for the more general problem of form discrimination will be considered.

The first of these hypotheses was that, in single-unit displays, the magnitude of a distance threshold for the recognition of symbols is directly related to the amount of information transmitted. The second hypothesis was that distance thresholds for multiple-unit displays are primarily a function of that unit which requires the most difficult discrimination. Implicit in this hypothesis is the assumption that multiple-unit displays involve independent and concurrent processes, each of which can be analyzed separately in terms of relationships established for single-unit displays. Data have been obtained which substantially confirm these hypotheses (**2, 3**).

In order to relate these findings to the problem of form discrimination, let us assume that a given set of forms can be considered as a multi-dimensional display composed of a number of characteristic features or dimensions, each of which conveys information. The process of form discrimination then reduces to one of discriminations within each of the constituent dimensions.

From the earlier reported findings, one can infer a procedure for predicting threshold values for a set of forms according to the difficulty of discriminating within each of the relevant dimensions. The magnitude of the threshold for the total set should then be equivalent to that for the dimensions which presented *S* with the most difficult discrimination.

This analogy between form discrimination and multiple-unit displays offers an explanation for the relationship often obtained between thresholds for the discrimination of forms and the amount of information transmitted by a choice. In practice, the coding of a larger number of alternatives usually involves either increasing the difficulty of discriminations within existing dimensions, or the addition of dimensions in which still more difficult discriminations are required. Under such circumstances, one would expect an increase in threshold values. Conversely, if the number of stimulus alternatives is decreased, it usually follows that the discrimination can be accomplished with fewer dimensions and with the elimination of the most difficult. Under such circumstances, one would expect response thresholds to decrease as the nature of the task is simplified.

However, there is no reason why the foregoing relationship must hold. One can design sets of forms such that an increase in the number of alternatives does not add more dif-

[1] The experiments summarized in this report were carried out while the author was a member of the Department of Psychology, Tufts University. The author wishes particularly to express his appreciation to Isaac Behar, presently at Emory University, for his assistance and cooperation in this work.

ficult discriminations. Here one would predict no increase in threshold even though the number of alternatives was increased. Similarly, one could reduce alternatives in such a fashion that the difficulty of discriminations was not decreased and thresholds for the remaining forms were not lowered.

This approach to the problem of form perception also has important consequences for clarifying the effect of *S*'s perceptual set on thresholds. In certain experiments, discrimination has been made easier not by simplifying the stimulus material, but by providing *S* with prior information for reducing the number of response alternatives (**1, 4, 5**). There remains the question of why restriction of response alternatives should simplify the nature of such tasks. The findings for multiple-unit displays when applied to form discrimination suggest a possible explanation.

Given a particular set of alternatives, restrictions on responses can be chosen so as to eliminate from the task the most difficult dimension of discrimination. One would then expect the threshold for the restricted set to be lowered and to be determined by the most difficult dimension remaining.

By a similar process of reasoning, one can predict what should happen if expected uncertainty exceeds actual uncertainty. Once *S* is oriented to expect a certain set of alternatives, he is prepared to discriminate all dimensions consistent with this induced orientation. If the expected set of alternatives implies a more difficult level of discrimination than is actually required, then one would expect threshold values to exceed those obtained when *S* has more accurate prior knowledge.

General Method

In order to test experimentally the validity of a multi-dimensional display model of form discrimination, a set of 16 forms was constructed (Fig. 1) and displayed in such a manner that a distance threshold for their discrimination could be determined. The set of forms was constructed from four binary choices, each of which constituted a single dimension. In selecting the constituent dimensions, an attempt was made to ensure that three levels of difficulty would be included. The three kinds of discrimination, in order of increasing difficulty, were as follows: (*a*) between the large square versus the large circle, (*b*) between the semicircle with the flat side up and that with the flat side down, and (*c*) between the small square versus the small solid circle at the geometric center of each large form. The actual dimensions of the elements making up these forms are as follows. The outside dimensions of the large square are $1\frac{1}{2}$ in. by $1\frac{1}{2}$ in. The width of the black border is $\frac{3}{16}$ in. The large circle was inscribed in the large square; the width of its border is also $\frac{3}{16}$ in. The smaller half-circles have a diameter of $\frac{3}{4}$ in. The small, solid-black square is $\frac{3}{16}$ in. on a side, while the small solid circle was inscribed in the small square. Thus the forms are composed of a set of black-on-white discriminations, the relative difficulty of which is varied primarily by manipulating the

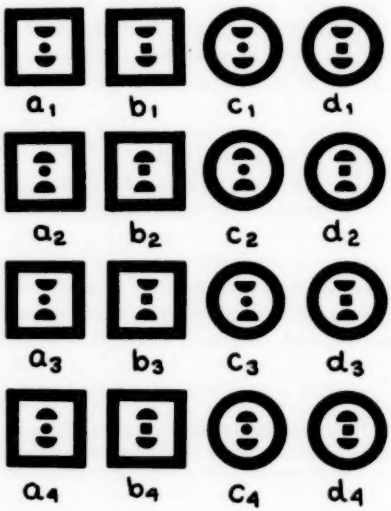

Fig. 1. Stimulus material used in all experiments.

area subtended by each of the dimensions at the retina.

The apparatus used for displaying these forms (**2**) consisted primarily of an enclosed box, the interior of which could be scanned by *S* through a fixed eyepiece equipped with a screen which decreased the clarity of the contours of forms viewed through it. Behind the screen was a movable carriage upon which a single form could be displayed. This carriage was driven by a constant-speed electric motor controlled by *S*. The movement of the forms during a trial was always toward *S* and along his line of sight. As the carriage advanced, the forms could be seen with increasing clarity until, ultimately, discriminations of considerable accuracy could be made. At the beginning of each trial, the display was positioned at a fixed starting point at such a distance that *S* could not distinguish the form. The threshold value was defined as the amount of travel required from the starting point to the point where recognition was possible. This distance was measured by means of a counter which recorded the movement of the drive shaft. Subsequently, these threshold values were converted into distance in inches from *S*'s eye (Tables 1–3).

After *E* had prepared the display for a given trial, *S* was instructed to position his head by means of the eyepiece, to start the movement of the carriage motor at his own convenience, and to continue without interruption. When he could recognize the form displayed, he was to release the control switch, thus stopping the motor, and to give his response verbally. Since none of the *S*s was familiar with this set of forms prior to the experiment, a procedure was devised for simplifying the manner in which *S*s reported their judgments. Each of the forms was given a symbolic designation as shown in Fig. 1. The *S*s were instructed to report in terms of this code rather than in terms of the actual form characteristics. In addition, the reporting procedure was organized to avoid requiring *S*s to memorize the code. During each experimental condition, they were provided with a cardboard display on which were mounted copies of each form possible during that condition with its appropriate code designation. The *S*s were instructed to identify the form they had just recognized on this auxiliary display and to report only the proper code designation. Instructions to *S* emphasized accuracy although he was also advised to give his response as soon as he was reasonably certain of his decision. During the practice trials which preceded each experimental period, full knowledge of the accuracy of responses was always given. Knowledge of threshold values or accuracy was not given to *S*s during the experimental trials. However, threshold values for inaccurate responses were always replaced by the value obtained from an additional presentation.

Experiment I

In this first experiment, threshold values were determined for the complete set of 16 forms and, separately, for each of the binary discriminations contained in the set. The hypothesis advanced was that the threshold for the discriminations of the total set would equal that obtained for the most difficult of the four binary discriminations.

Method.—Although the total discrimination involves four binary choices, these represent only three levels of difficulty. Thus, in addition to thresholds for the total discrimination, values were also determined for the large square versus large circle, for the semicircles facing either up or down, and for the small square versus small circle. Assuming that the latter discrimination is the most difficult of the three, it was postulated that the threshold value for that discrimination would determine the value for the total set. In order to obtain threshold values for these binary discriminations, each pair was displayed separately.

Each *S* discriminated forms under all of the four experimental conditions during a single experimental period. Each period was begun with a series of practice trials designed to familiarize *S* with the binary choices contained in the forms and with all of the discriminations required. The *S* was first given practice in discriminating each of the binary dimensions. These were presented in order of increasing difficulty. Four practice trials were administered for each of the elementary discriminations. The *S* was then shown the set of 16 forms and shown how they were constructed from the elementary dimensions. Six practice trials were then administered on this total set except in a few instances where additional practice presentations were needed before, in *E*'s judgment, *S* was prepared for the experimental procedure. Finally, the four experimental conditions were administered. Under Cond. 1, *S*s discriminated between the large square and large circle; under Cond. 2, between the pair of semicircles; under Cond. 3, between the small square and small circle; and under Cond. 4, among the total set of 16 forms. Six trials were given for each of the first three conditions and 12 for the final condition.

TABLE 1

VARIATIONS IN THRESHOLD AS A FUNCTION OF THE CODING OF THE INFORMATION

Cond.	Mean	*SD*	Paired Diff.		*t*
1	40.16	.346			
2	38.80	.182	1–2	1.36	16.35*
3	38.48	.125	2–3	.32	8.39*
4	38.49	.103	3–4	.01	.73

* $P = .01$, $df = 7$.

Eight *S*s were used and rotation of conditions was introduced in order to compensate for possible practice effects. For each *S*, a sequence of presentations within each condition was chosen randomly with replacement. Moreover, all *S*s operated within a given condition with full knowledge of the set of possible alternatives.

Results.—The results are summarized in Table 1. Throughout the analysis, the average of the scores for all presentations within a given condition was computed for each *S* under each of the experimental conditions. The *t* test was applied to the paired differences. For those hypotheses in which the direction of the paired differences was specified initially, one-tailed tests of significance were employed. With respect to the comparison of Cond. 3 with Cond. 4, for which the direction of the difference had not been predicted, a two-tailed test of significance was employed. The same statistical procedures were followed in the other experiments reported here.

From these data, the relative difficulty of the three binary discriminations may be compared. As would be expected, the small square versus small circle is the most difficult of the three dimensions. The hypothesis that the difficulty of the total set is determined by the difficulty of Cond. 3 (the most difficult of the series of elementary discriminations) may be examined. As was predicted, there is no significant difference between the thresholds for these two discriminations.

EXPERIMENT II

The effect on threshold values of varying *S*'s perceptual set was investigated in Exp. II. In the first half of this experiment (Part I) *S*s were instructed to ignore the dimension containing the most difficult discrimination. It was postulated that such a restriction on response alternatives would lower thresholds, and that the resulting threshold would be determined by the most difficult discrimination remaining.

In the second half (Part II), attention was focused on the consequence of expected uncertainty exceeding actual uncertainty. For this portion of the experiment, *S* was led to expect a set of alternatives which included a dimension requiring a more difficult discrimination than was actually necessary. The prediction made was that *S*'s task would be unnecessarily complicated by this inefficient orientation, and that his thresholds would be determined by the most difficult level of discrimination expected rather than the most difficult discrimination actually present.

Method.—For Part I, *S*s were asked to discriminate forms under three sets of conditions. For Cond. 1, they were presented with the total set of forms and were asked to report in terms of all the possibilities. For Cond. 2, they were again presented with the total set but were instructed to ignore the most difficult dimension, i.e., to refrain from discriminating the small square from the small circle. For Cond. 3, they were asked to discriminate between the two semicircles, since it was this elementary discrimination which was expected to determine the threshold.

For Part II, *S*s were again asked to discriminate forms under three sets of conditions. For Cond. 4, they were asked to discriminate the total set of forms from which the presenta-

TABLE 2

VARIATIONS IN THRESHOLD UNDER RESTRICTED KNOWLEDGE OF THE APPROPRIATE SET

Cond.	Mean	*SD*	Paired Diff.		*t*
1	38.53	.087			
2	38.76	.104	2–1	.23	23.34*
3	38.82	.162	3–2	.06	1.72
4	38.52	.048			
5	38.57	.101	5–4	.05	1.45
6	38.48	.096	6–5	−.09	1.85

* $P = .01$, $df = 5$.

tions were chosen. For Cond. 5, instructions were simply to discriminate "some more of the complex forms." However, for this condition, presentations were restricted to the eight forms containing the small circle (Categories a_1 through a_4, c_1 through c_4). This restriction was not made known to *Ss*.[2] For Cond. 6, the choice between the small square and small circle was presented, since it was this discrimination which was expected to determine the difficulty for both of the more complex conditions. Each half of this experiment was administered to six *Ss* in a single experimental period, and was conducted in the same way as Exp. I.

Results.—The results are summarized in Table 2. The first set of results summarizes the effect of response restriction on task difficulty. Comparison of the results for Cond. 1 and 2 indicates that the instruction to ignore the most difficult dimension significantly simplifies the task. In addition, it may be noted that the threshold for Cond. 3 is not significantly different from that obtained for Cond. 2, indicating that when response alternatives are restricted, the difficulty of the task is primarily determined by the most difficult dimension that remains.

The results of Part II demonstrate the consequence of expected uncertainty exceeding actual uncertainty. The results for Cond. 4, 5, and 6 are statistically indistinguishable one from another. One can draw the inference that *Ss* under Cond. 5 continued to behave as if the dimension of small square versus small circle were still the controlling element in determining the threshold for their responses.

[2] Apparently, no *Ss* were aware of this restriction although one *S* commented at the conclusion that this particular condition had contained an unusually large number of "small circles" among the presentations.

EXPERIMENT III

In Exp. II it was demonstrated that, under certain specified conditions, a restriction on the number of response alternatives leads to a simplification of the task of form discrimination. In a final experiment, an attempt was made to demonstrate that the result obtained may be expected only if the restriction on response alternatives permits *S* to ignore the most difficult dimension of the total task.

Method.—In Part I of Exp. III, a subset of the total set of alternatives was chosen which eliminated the most difficult discrimination. It was hypothesized that such a restriction on response alternatives would lead to a reduction in *S*'s threshold. The *Ss* discriminated forms under three sets of conditions. For Cond. 1, they were presented with the total set of forms and were asked to report in terms of the complete set of possibilities. For Cond. 2, stimulus alternatives were limited to Subsets a_1 through a_4, b_1 through b_4, and c_1 through c_4. This restriction was known to *S*. In addition, special instructions were given in order to limit the response alternatives open to *S*. He was instructed to report all forms containing a large square as a single category, and was told that it was unnecessary to discriminate within this subset, designated as Cond. 2a. However, for the response subset of "large circles," designated as Cond. 2b, he was asked to identify the specific form being displayed. The effect of these instructions is that the discrimination between small square and small circle is unnecessary for responses in Cond. 2a and 2b, while the discrimination between semicircles is required only in Cond. 2b. For Cond. 3, all 16 stimulus alternatives were possible and this fact was known to *S*. Again *S* was in-

structed to treat the subset of "large squares," designated as Cond. 3a, as a single response category but to discriminate completely within the subset of "large circles," designated as Cond. 3b. With this set of response instructions, the thresholds for Cond. 3a should be lowered, but not for Cond. 3b in which all levels of difficulty are still required for discrimination.

In Part II of Exp. III, a response restriction was chosen which retained that dimension containing the most difficult discrimination. It was hypothesized that no reduction in *S*'s thresholds would result. As before, *S*s discriminated forms under three sets of conditions. With respect to both stimulus alternatives and response instructions, Cond. 4 is identical with Cond. 1 of Part I of Exp. III, while Cond. 6 is identical with Cond. 3. For Cond. 5, stimulus alternatives were limited to Subsets a_1 through a_4, b_1 through b_4, c_1 and c_2, d_1 and d_2. This restriction was known to *S*. In addition, response alternatives were limited by instructing *S* to treat the subset of "large squares," designated as Cond. 5a, as a single category, but to discriminate all forms within the subset of "large circles," designated as Cond. 5b. These instructions permitted *S* to eliminate the discrimination between small square and small circle only for Cond. 5a. For Cond. 5b all levels of difficulty involved in the total discrimination task were still required and no reduction in threshold was anticipated.

Each half of this experiment was administered to six *S*s in the same manner as for Exp II, with the exception of the method for choosing the random sequences of presentations for those conditions with response instructions involving "large square" and "large circle" subsets (Cond. 2, 3, 5, and 6). For each condition, 12 presentations were administered. For Cond. 2, 3, 5, and 6, half of the presentations were chosen randomly with replacement from each of the appropriate response subsets.

Results.—The results are summarized in Table 3. Considering the results for Part I, the thresholds for Cond. 2a and 3a are significantly lower than those for any other conditions. This is consistent with the fact that neither of the two more difficult levels of discrimination were required. In addition, the threshold for Cond. 2b is significantly lower than that obtained for Cond. 1, although the threshold for Cond. 3b is not. This latter result is consistent with the fact that Cond. 3b still involves the most complex dimension of discrimination, which was eliminated from Cond. 2b. Similar results were obtained from Part II of the experiment. As before, the thresholds for Cond. 5a and 6a are significantly lower than those obtained for any remaining conditions. The threshold for Cond. 6b, which was defined to include the most difficult discrimination, is not significantly different from that obtained for Cond. 4. The results for Cond. 5b are not consistent with the initial prediction. This subset also involved the most difficult level of discrimination, yet its threshold was slightly lower than that for Cond. 4. Since the difference is small in absolute value and all other predictions were confirmed, this result does not seem to invalidate the hypotheses being investigated.

TABLE 3

VARIATIONS IN THRESHOLD AS A FUNCTION OF INDUCED SET

Cond.	Mean	*SD*	Paired Diff.		*t*
1	38.69	.096	2a–2b	1.58	23.41**
2a	40.51	.163	3a–3b	1.73	45.50**
2b	38.93	.113	2b–1	.24	6.24**
3a	40.44	.089	3b–1	.02	1.75
3b	38.71	.090	2b–3b	.22	6.95**
			3a–2a	−.07	1.07
4	38.71	.126	5a–5b	1.69	24.22**
5a	40.44	.121	6a–6b	1.72	41.51**
5b	38.75	.052	6b–4	.03	1.64
6a	40.46	.253	5b–4	.04	2.69*
6b	38.74	.162	5b–6b	.01	.10
			6a–5a	.02	.26

* $P = .05$, $df = 5$.
** $P = .01$, $df = 5$.

SUMMARY

Three experiments have been reported on the determination of distance thresholds for discrimination of a specially constructed set of forms. In Exp. I the hypothesis was advanced that the magnitude of the threshold for the

total set would be equivalent to the value obtained for that dimension which presented *S* with the most difficult discrimination. The results obtained were consistent with this hypothesis.

Experiment II was focused upon the effect on threshold values of altering the perceptual set of *S*s toward the possibilities being presented. In Part I of this experiment, *S*s were instructed to ignore some of the information contained in the presentations. This restriction on responses was chosen so that *S*s could ignore the most difficult dimension of the total discrimination. The prediction was that the threshold for the revised set of alternatives would be determined by the threshold for the most difficult dimension remaining. The results confirm the hypothesis. In Part II of this experiment, the effect of expected uncertainty exceeding actual uncertainty was investigated. The results indicate that thresholds obtained under such conditions exceed those obtained when *S* has full knowledge of the nature of the discrimination task.

In a final experiment, an attempt was made to demonstrate that restrictions on the number of response alternatives simplify the task of form discrimination only if this restriction enables *S* to ignore that dimension containing the most difficult discriminations. With one exception, the results obtained were consistent with the hypothesis.

References

1. Frick, F. C. Some perceptual problems from the point of view of information theory. In R. A. Patton, *Current trends in information theory.* Pittsburgh: Univer. of Pittsburgh Press, 1953, pp. 76–91.
2. Krulee, G. K., Podell, J. E., & Ronco, P. G. Effect of number of alternatives and set on the visual discrimination of numerals. *J. exp. Psychol.*, 1954, **48,** 75–80.
3. Krulee, G. K., & Weisz, A. Studies in the visual discrimination of multiple-unit displays. *J. exp. Psychol.*, 1955, **50,** 316–324.
4. Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the text material. *J. exp. Psychol.*, 1951, **41,** 329–335.
5. Postman, L., & Bruner, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39,** 359–377.

(Received February 15, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

ASSIMILATION AND CONTRAST EFFECTS OF ANCHORING STIMULI ON JUDGMENTS[1]

MUZAFER SHERIF, DANIEL TAUB
University of Oklahoma
AND
CARL I. HOVLAND
Yale University

In complex social communication the introduction of reference points may produce two opposing effects. Under some conditions the introduction of a reference point or stand beyond *S*'s current position tends to move him toward the new position. Thus, telling him that experts think it will be at least 10 yr. before peaceful use of atomic power is feasible, may cause an individual to increase his own estimate from one of 5 yr. to one of 6 or 7 yr. Under other conditions the introduction of communication results in a rejection of the new proposal and a stronger entrenchment in his original position. Here one has the frequently mentioned "boomerang effect" (**2, 3, 7, 9**). To some extent at least, these phenomena may be the result of judgmental processes and may be conceptually closely akin to the phenomena of assimilation and contrast in the judgment of simple stimulus material. Before proceeding to a study of the complex factors involved in communication it was thought desirable to start with an analysis of these judgmental effects with simple stimulus materials in a manner applicable to the social area.

The research of Rogers (**11**) suggested the feasibility of doing such a study with weight-lifting. Accordingly, the procedures closely paralleled his. Judgments were first obtained as to the categorization of stimuli from 55 to 141 gm. in weight along a 6-point scale, using the method of single stimuli. In the second and subsequent sessions the weights were presented in pairs with the first ("anchor") weight described as being in the topmost category. The anchoring stimuli ranged from 141 to 347 gm. While the procedure was essentially a replication of Rogers' experiment, his formulation was not in the present terms and unfortunately he does not present his data in a form permitting the type of analysis required for the present hypotheses.

The general hypothesis of the study is that whether one obtains an "assimilation effect" or a "contrast effect" depends on the position within or distance from the original series of the introduced anchors (standards). When an anchor is introduced at the end or slightly removed from the end of the series, there will be a displacement of the scale of judgment toward the anchor and assimilation of the new reference point in the series. When, however, the reference point is too remote there will be displacement in the opposite direction (i.e., away from the anchor), with a constriction of the scale to a narrower range.

[1] This study was conducted as part of the Yale Communication and Attitude Program which is supported by a grant to Carl I. Hovland from the Rockefeller Foundation. The Foundation's support is gratefully acknowledged.

More formally, two hypotheses are presented for testing:

1. In judgments of graded stimuli ranging from low to high in some dimension, the introduction of anchors at the end points of the series or immediately above or below the series will cause displacement in the distribution of judgments of series stimuli in the direction of the anchor (assimilation effect).

2. As the anchors are placed at increasing distances from the upper or lower ends of the series, the distribution of judgments will be displaced in the direction away from the anchor and the whole judgmental scale will be constricted (contrast effect).

One or the other of these effects has been suggested by a number of previous experiments (**1, 4, 5, 10, 11, 13**) but the factor of distance has not been systematically varied to investigate both phenomena in a single study.

Procedure

The basic procedure consisted of obtaining judgments of weight by the method of single stimuli with and without anchoring stimuli. Judgments were made in terms of a 6-point scale, "1" being the lightest, "6" the heaviest. The anchoring stimuli in the main experiment were all equal to or greater than the top stimulus value in the original stimulus series and each was described as a stimulus within the topmost category.

The apparatus consisted of a frame designed to offer sturdy support for the mechanical lifting of a set of weights. The lifting system consisted of light cordage, heavily waxed, placed horizontally over two plastic slide wheels. A horizontal pull upon a handle attached to the cordage provided for an upward motion of the weight attached. Since weight containers were of equal height and the cordage was knotted, pull was of uniform length for the different weights.

The weights were concealed from *Ss* by a masonite board pierced appropriately to permit passage of the cordage. This procedure of weighing provided few cutaneous cues to *S* and proved to be a satisfactory procedure for the study. Pretests established that most *Ss* could comfortably place the stimuli in the series (without introduced anchors) correctly at least 50% of the trials.

The stimulus series consisted of weights of 55, 75, 93, 109, 125 and 141 gm. and is referred to here as the "original stimulus series." The anchoring stimuli were weights of 141, 168, 193, 219, 244, 267, 288, 312 and 347 gm. It will be noted that the 141-gm. anchor is the same as the topmost weight in the original series.

In the first session, the operation of the mechanism was demonstrated by *E*. Then the following instructions were given: "The stimuli you are to judge will be a series of weights. You are to lift these by pulling up the handle as demonstrated. Following the lifting of a weight you are to judge its location on a scale of 1 to 6 inclusive. Make a judgment after each lifting even though at first you may not have adequate basis for judgment."

The weights were presented in haphazard order, one by one in sections of 50 presentations with a rest period of 5-10 min. between sections.

Only those *Ss* who correctly discriminated the original stimulus series in at least 50% of the trials were used in subsequent sessions in which anchors were introduced. Two *Ss* were not used because they failed to meet this criterion.

In anchor sessions the following instructions were read: "The stimuli will now be presented in pairs instead of singly. The first member of each pair will tell you what you are to call 6, the second you are to judge, as before, on a scale of 1 to 6. Tell me of your judgment after the second member of each pair even though at first you may not have an adequate basis for judgment."

These instructions were read to *S* at the beginning of each of the nine anchor sessions. The order in which anchors were introduced in successive sessions was haphazard. Only one anchor weight was used in any one session. No *S* took part in more than one session on any one day.

Subjects.—The *Ss* in the main experiment were six male students from the University of Oklahoma. They were naive *Ss* who did not know the purpose of the experiment and had not previously served as *Ss* in a psychological experiment. Each *S* gave 300 judgments per session and each served in 10 experimental sessions (one with only the original stimulus series and one with each of the nine anchors). Thus, 3000 judgments were obtained from each *S*.

In addition to the above, extensive pretesting was carried out. In this pretesting phase the

anchor beyond the scale was removed far beyond the highest anchor (viz., 347 gm.) reported in this paper.

Results

The principal results of the experiment are shown in Fig. 1, where the distribution of judgments of weights is given for the original stimulus series (without anchor) and for the same stimuli when judged with anchors at, or varying distances above, the top stimulus in the series. It will be observed that the distribution of judgments in the original series without anchor is rectangular and does not deviate too greatly from the flat distribution represented by equal frequencies in each category. When an anchor, whose weight in grams corresponds to the top stimulus value in the original stimulus series is used, the result is a displacement of the distribution in the direction

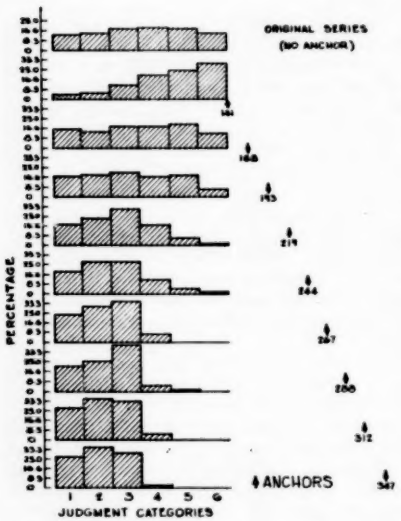

Fig. 1. Distribution of judgments for original series of weights without anchor (top) and with anchors at increasing distances above original series (heaviest anchor at bottom).

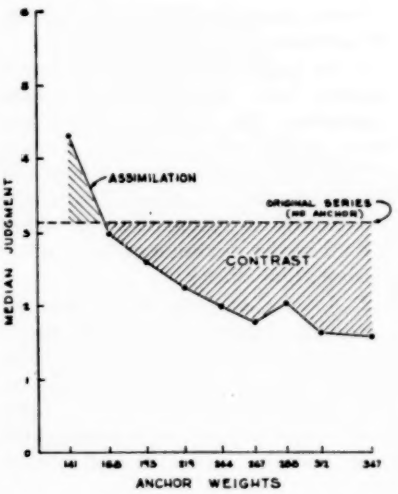

Fig. 2. Median judgments with anchors at increasing distances from the series (abscissa) plotted against the median judgment without anchor.

of the anchor. Anchors more distant from the top stimulus value in the series tend to produce the opposite effect, causing a displacement of judgments to the lower end of the scale. The effects shown in Fig. 2 hold for each *S* with only minor variation. These effects can be represented more clearly by plotting the median judgment given to the stimulus series without anchor and with anchors varying distances from the original stimulus series, as in Fig. 2. The median judgment during the original series without anchor is used as the baseline from which to evaluate change.

Supplementary Experiments

The preceding results indicate the operation of the predicted assimilation and contrast when the anchor is above the series. It seemed desirable to replicate the experiment

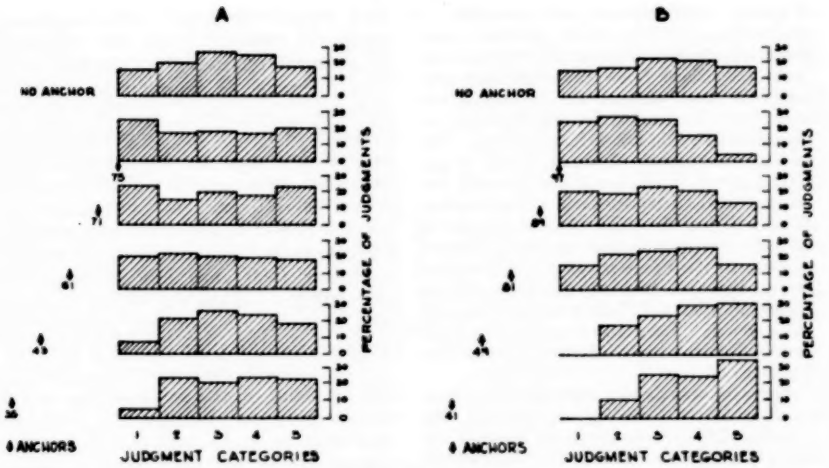

FIG. 3. Distribution of judgments for two different stimulus series without anchor and with anchors at distances below each series (indicated by arrows).

to determine whether the effects obtained in the main experiment are also found when anchors are introduced at the lower end of the original stimulus series and at increasing distances from it. This was demonstrated in a supplementary experiment which was carried out in two separate series.

Series I [2]

Procedure.—The stimulus series used in the main experiment had to be modified because the remaining range to zero point below the lower end of the original series made the introduction of only a few anchors possible. Accordingly the lightest weight (55 gm.) was eliminated and the series of five remaining weights was used (75, 93, 109, 125 and 141 gm.). Anchors of 75, 71, 67 and 43 and 35 gm. were employed. It will be noted that the first three anchors were close to the series and actually within the original stimulus series used in the main experiment. They were chosen to be maximally sensitive to assimilation effects. The lighter anchors were selected as conducive to contrast with a gap in the region between 67 and 43 gm.

Instructions to *Ss* were substantially the same as those in the main experiment. The principal procedural difference involved the use of direct lifting of weights in cylindrical containers 9.6 cm. high and 5.5 cm in diameter.

Each *S* served for six sessions (one original series session without anchor and five anchor sessions) with from 12 hr. to 3 days between sessions. Only one anchor was used during a session. Four different orders of anchors were used for the 4 *Ss*.

Results.—The distributions without anchor and with anchors at varying distances from the series are shown in Fig. 3A. It will be noted that both the 75-gm. anchor and the 71 gm. anchor, which was slightly below the end stimulus, produced shifts in the distribution toward the anchor ("assimilation"). Anchors further removed from the end produced a contrast effect similar to that shown in Fig. 1 although the effects are less pronounced.

Series II [3]

Procedure.—Use of the original stimulus series in the main experiment and introduction

[2] Series I was conducted by B. Jack White, University of Oklahoma.

[3] Series II was conducted by William R. Hood, University of Oklahoma.

of anchors below the series, even as used in the modified procedure in Series I above, posed procedural problems, since too few possible anchors remained below the series and since sensitivity and judgment are atypical in this range (**14**). So a new series of five weights was prepared of 97, 118, 137, 153, and 169 gm. The anchors selected were 97, 89, 81, 49 and 41 gm.

In pretests, it was found that *S*s could correctly place the five test stimuli without anchors at least 50% of the trials. Special note should be made that the highest anchor is identical in value to the lowest stimulus in the original series. Intervals between the three highest anchors were less than half the distance between series stimuli. There was a gap between these three and the two anchors further removed from the series.

Instructions to *S*s were substantially the same as those in the preceding experiments. The principal difference in procedure involved the direct lifting of weights, which were in cylindrical containers 8.3 cm. high and 8.5 cm. in diameter.

Each *S* served for six sessions (one for the original series and one for each anchor) with from 12 hr. to 2 days between sessions. Only one anchor was used during a session. Three different orders of anchors were used for the 3 *S*s.

Results.—In Fig. 3B the distributions of judgments for the five test weights without anchor and with anchors varying distances removed are presented. It appears quite likely that an assimilation effect occurs in the lower range with the first anchor and the anchor just removed from the series. Anchors further removed resulted in contrast effects much like those obtained in the main experiment. All three *S*s show the assimilation and contrast effects indicated for the average.

Discussion

The foregoing results support our hypothesis in showing a range immediately above or below the stimulus series in which assimilation occurs (*assimilation range*) and another range beyond this where contrast occurs (*contrast range*). In the present experiment the assimilation range is rather narrow. We would expect that the range of assimilation would be wider with a greater range of stimuli. On the basis of judgmental relativity we would expect that the range of assimilation would be influenced not only by the absolute distance of the anchor from the extreme stimuli but also by the range of the stimulus series itself. Thus a series from 500–700 gm. would undoubtedly have different assimilation and contrast ranges than a series with the same midpoint (600 gm.) but ranging from 400–800 gm. Thus the effects of scales and anchors are not independent but are reciprocal in their effects.

The present demonstration helps to clarify some of the relationships which obtain in social judgments. Here we find that an individual places a verbal statement on an issue both in terms of the item's relative proximity to his own position and the latitude which is acceptable to him around that focal point of acceptance (**8**). This result is not easily explained in terms of "indifference points" or "neutrality regions" (**5, 6**). Similarly, in an earlier study (**12**) we found that a person's position on a social issue serves as the major anchor when he places opinion statements concerning the issue into what he considers the most appropriate number of different categories. The *S*s with extreme stands on the issue tended to use very few categories and to utilize some categories with disproportionate frequency, while *S*s with middle-of-the-road positions distributed their judgments into a larger number of categories. Here again the displacement of judgments was related to *S*'s position on the issue, which served as a major anchor.

Summary

The hypothesis was tested that introduction of anchoring stimuli immediately adjacent to the stimuli being judged will cause displacement of judgments in the direction of the anchor. This would constitute an "assimilation effect." Anchors considerably beyond the stimulus range were also employed to replicate previous findings that this will produce a displacement of judgments away from the anchor ("contrast effect").

In the first experiment weights of 55, 75, 93, 109, 125 and 141 gm. were judged. The anchoring stimuli were weights of 141, 168, 193, 219, 244, 267, 288, 312 and 347 gm. In the first session judgments were made without anchors. In subsequent sessions anchors were introduced before each stimulus being judged. Six male college students gave 300 judgments per session. Clear evidence for the hypothesis was obtained, with pronounced displacement of judgments *toward* the anchor when it was introduced at the top of the series and displacement *away* with more distant anchors.

In two further experiments, anchors were introduced at and below the lightest weights in the series. Again anchors at or slightly removed from the stimulus series resulted in displacement of judgments toward the anchor, while anchors more distant from the series produced the familiar displacement away from the anchor.

The relationship of these results to allied phenomena with social judgments is discussed.

References

1. Beebe-Center, J. G. *Pleasantness and unpleasantness.* New York: Van Nostrand, 1932.
2. Cooper, E., & Jahoda, M. The evasion of propaganda: how prejudiced people respond to anti-prejudice propaganda. *J. Psychol.*, 1947, **23,** 15–25.
3. Flowerman, S. H. The use of propaganda to reduce prejudice: a refutation. *Internat. J. Opinion & Attit. Res.*, 1949, **3,** 99–108.
4. Heintz, R. K. The effect of remote anchoring points upon the judgment of lifted weights. *J. exp. Psychol.*, 1950, **40,** 584–591.
5. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60,** 1–29.
6. Helson, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55,** 297–313.
7. Hovland, C. I., Janis, I. L., & Kelley, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.
8. Hovland, C. I., & Sherif, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55,** 244–252.
9. Lazarsfeld, P. F. Communication research and the social psychologist. In W. Dennis, (Ed.), *Current trends in social psychology.* Pittsburgh: Univer. of Pittsburgh Press, 1948, 218–273.
10. Long, L. A study of the effect of preceding stimuli upon the judgment of auditory intensities. *Arch. Psychol.*, 1937, No. 209.
11. Rogers, S. The anchoring of absolute judgments. *Arch. Psychol.*, 1941, No. 261.
12. Sherif, M., & Hovland, C. I. Judgmental phenomena and scales of attitude measurement: placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48,** 135–141.
13. Volkmann, J. The anchoring of absolute scales. *Psychol. Bull.*, 1936, **33,** 742–743. (Abstract)
14. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received February 25, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

ACHIEVEMENT DRIVE AND HABITUAL MODES OF TASK APPROACH AS FACTORS IN SKILL TRANSFER [1]

GUY H. MILES [2]

State University of Iowa

The analysis of proactive and retroactive practice effects has been the aim of several recent studies of complex perceptual-motor skills. In these studies *Ss* first practice a complex motor task during an original learning (OL) phase. Next, during an interpolated learning (IL) phase, they practice on a task which is *basically* the same but requires different directional movements of the controls. Finally, *Ss* return to the original task in a relearning (RL) phase. Performance gains and/or losses during IL which occur as a result of first practicing the OL task are called proactive effects; gains and/or losses during RL occurring as a result of practice on the IL task, retroactive effects.

The *Ss* practicing the OL, IL, and RL tasks under uniform conditions differ markedly from each other in proficiency at the outset of both IL and RL. These performance differences cannot reasonably be attributed to differences in ability to perform the OL task. Several studies (**6, 7, 8, 9, 15**) that utilized different complex tasks have demonstrated that measures of performance on the last OL trial are correlated at quite low levels with performance measures on either the first IL or the first RL trials.

[1] This investigation was made during the author's tenure as a Research Fellow of the National Institute of Mental Health. It was supported in part by National Science Foundation Grant G635, and was supervised by Professor Don Lewis.

[2] Now at University of Minnesota.

Attempts have been made to explain these performance differences among individuals in terms of alleged personality traits (**2, 3, 14**). The existence of such traits has not been substantiated by empirical evidence (**7**). Individuals who show extreme proactive or retroactive performance losses in one performance situation fail to do so in another.

The present study is an attempt to discover some of the determinants of these differences among individuals in performance on a single perceptual-motor device: the Iowa Pursuitmeter.

Observations of *Ss* during their practice on various motor tasks led to the belief that different individuals approach tasks in different ways. Some *Ss* appear to make haphazard approaches, depending largely on trial-and-error behavior to solve the manipulatory problems involved. Others apparently try to "figure out" the underlying relationships and requirements. Some *Ss* move slowly and deliberately, others rapidly with no apparent plan.

It was conjectured that these behavioral differences reflect long established habits of approaching novel, complex tasks. Any such habit might be facilitative of performance on one task but not on another; might aid in the recognition of the underlying features of one task but not of another. Because of the role of task similarity in transfer, it was supposed that an approach facilitative of learning certain features common to both the OL and IL

versions of a task would result in more positive transfer than would an approach not facilitative of learning these task features.

Drive level would be expected to interact with habitual modes of approach, to produce OL performance level differences. If the approach were facilitative of learning the features common to both the OL and IL task versions, drive would energize the habit strengths acquired during this learning and produce further differences in proactive and retroactive effects.

The present study is an attempt to measure the effects and interactions of achievement motivation and *S*s' tendency to approach complex tasks analytically.

Method

Manipulation of drive and analytic variables.—Since the Block Design subtest of the Wechsler Bellevue Intelligence Scale is widely used as a measure of ability to analyze and synthesize, an exploratory study was made to determine whether scores on the Block Design test correlated with performance measures on the Iowa Pursuitmeter. It was found that Block Design scores as such were *not* related to pursuitmeter performance. However, when *S*s were carefully questioned as to how they went about doing the designs, it was found that a high (that is, short time) score did not necessarily reflect an analytical approach. Those *S*s who indicated verbally that they had divided the Block Design pictures into parts before moving the blocks into position performed at a higher level on the pursuitmeter and showed less retroactive performance loss than did those indicating some other approach.

In the present study *S*s were divided into two groups, analyzers and non-analyzers, in the following way: block designs 2, 6, 8, 9, and 10 from the Wechsler Adult Intelligence Scale were presented singly in a standard sequence. If *S* was unable to complete a design within 10 min., he was presented the next one.

At the completion of the last design *S* was told, "There is no right or wrong way to do these designs, but we are interested in finding out how people make the designs. Would you tell me how *you* go about making them?"

If *S* gave a clear statement indicating that he had conceptually divided the several pictures into separate blocks, he was classified as an analyzer. He was also classified as an analyzer if, with his finger, he marked off the blocks on the design while saying, in effect, "I knew this block had to be half white and half red so . . . ." However, if *S* said he "just twisted the blocks" or "tried the blocks until they looked right" he was classified as a non-analyzer. Any *S* who was observed attempting to make a design by using more than three blocks on a side was classified as a non-analyzer. If *S* made an uninterpretable statement, he was asked to "tell me more about it" until his answer provided a basis for classification.

To introduce drive as a variable, it was decided to employ two groups of *S*s, selected on the basis of their achievement imagery (AI) scores on the Iowa Picture Interpretation Test (IPIT), a multiple choice thematic apperception test (**4, 5**). The assumption was that the groups differed markedly in drive level. Farber (**4**) points out that although variations in performance among groups that differ in achievement motivation can possibly be explained in terms of differences among the groups in associative factors, the extensiveness of the range of responses apparently affected by the achievement variable lends some credibility to the supposition that achievement imagery scores reflect strength of drive. Hurley (**5**) using the IPIT, found the effects of achievement motivation in verbal learning situations to be consistent with predictions made on the assumption that IPIT scores reflect degree of drive.

Subjects.—The *S*s were 68 male volunteers from classes in elementary psychology. The AI scores of half of the *S*s were in the top 25% of such scores for the combined classes, and of the other half, in the bottom 25%. Equal numbers of *S*s in each drive level group had been classified as analyzers and non-analyzers. This resulted in four experimental groups of 17 *S*s each. These groups are identified as HI-A—high drive analyzers, HI-NA—high drive non-analyzers, LO-A—low drive analyzers, and LO-NA—low drive non-analyzers. As a further control for drive level, the four groups were equated for Taylor Manifest Anxiety Scale scores (**16**).

Apparatus.—The tasks were provided by a modified form of the Iowa Pursuitmeter, described elsewhere (**1, 6, 11**). The *S* makes responses basically like the ordinary ones of steering and pointing in his efforts to keep a spot of light on a moving target.

The target, a 1-in. square area painted in the center of a flat surface 11 in. square, is

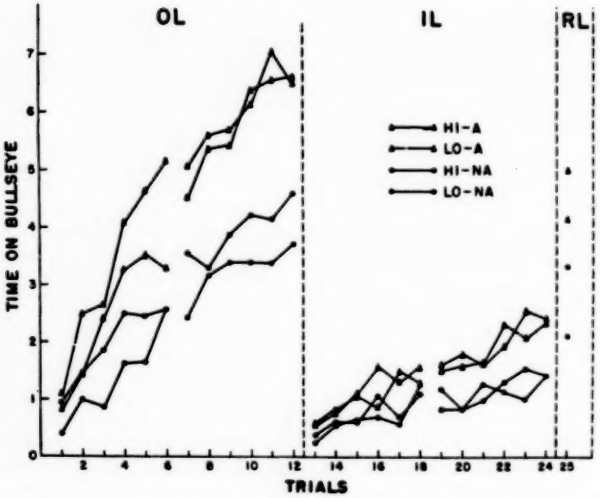

FIG. 1. Performance curves for four groups of male *Ss* (HI-A = high drive analyzers, HI-NA = high drive non-analyzers, LO-A = low drive analyzers, and LO-NA = low drive non-analyzers) during OL, IL, and RL practice on the Iowa Pursuitmeter. Based on means of time on bullseye per trial.

located at the end of a boom, about 5 ft. from *S*'s face. It moves through the same irregular pathway during each 30-sec. trial period, being returned automatically to the starting point during the rest interval between trials. The position of the spot of light which *S* strives to keep on target, is controlled through movements of the response unit.

The *S*, sitting on a chair, grasps two pistol-grip type handles placed at about chest height. During practice on the standard task, in order to move the light to the right or left, the handles and the movable structure to which they are attached are turned through a small angle by pushing one handle away while pulling the other toward the body in a steering motion. To make the light go up, the bottoms of the handles are pushed a short distance away from *S* in a kind of twisting action, and to make it go down, the handles are pulled toward *S*. The reversed task requires that *S* make responses exactly opposite to those required by the standard task. The speed with which the light moves in any direction depends on how far the handles are turned from the neutral position.[3]

Procedure.—Each *S* was given a total of 25 trials on the pursuitmeter. Each trial was 30 sec. in length, and the trials were separated by 30-sec. rest intervals. Longer rests of 1.5 min. were given after blocks of 6 trials and/or after the completion of practice on a task. All *Ss* began with 12 OL trials on the standard task, followed by 12 IL trials on the reversed task. They then had 1 RL trial on the standard task.

Two response measures were used: time on the 1-in. square, reported as "time on bullseye," and time on the 11-in. square (including time on the 1-in. square) reported as "time on large target."

RESULTS

The data for time on bullseye are summarized in Fig. 1, where means in seconds are plotted against trials. During OL the highest over-all performance level is shown for the high-drive analyzers. The lowest performance level, and the slowest rate of acquisition, are for the high-drive non-analyzers.

A trend analysis [Lindquist (**10**), Type I design] of the data over Trials 7–12 of OL revealed that the general means of performance are significantly

[3] See the paper by Miles and Lewis (**11**) for a schematic drawing of the Iowa Pursuitmeter.

different at better than the .5% level, and the trials means at better than the .1% level. The slopes of the curves do not differ significantly. Separate *t* tests indicated that the difference between the general means of the HI-A and HI-NA, and the LO-A and HI-NA groups differ from each other at better than the 1% level, while the HI-A and LO-NA group means are significantly different at the 2% level, and the LO-A and LO-NA group means at the 5% level. The differences between the means of the LO-A and HI-A, and the LO-NA and HI-NA groups are not statistically dependable.

No analysis was made of the IL bullseye data since the scores for time on large target seemed more useful in view of the large number of zeros among the bullseye scores.

The data for time on large target are summarized in Fig. 2, where means in seconds are plotted against trials. This response measure does not differentiate the groups during OL as well as does time on bullseye. However, the HI-NA group is again the lowest of the four groups throughout OL. During IL, the response measure is more revealing. The curves for IL indicate that the LO-A group performed at a higher level than the other groups. Evidence for a shift in the relative performance levels, and by inference a difference in proactive effects, of the HI-A and LO-A groups is indicated by a reversal of their relative performance levels on Trial 1 of OL and Trial 1 of IL. There is no clear evidence of such a shift in the relative performance levels of the other two groups.

A trend analysis of the time on large target data for all IL trials revealed that the general means of performance differ significantly at almost the 1% level ($F = 4.11$, 3 and 64 *df*). Separate *t* tests indicate

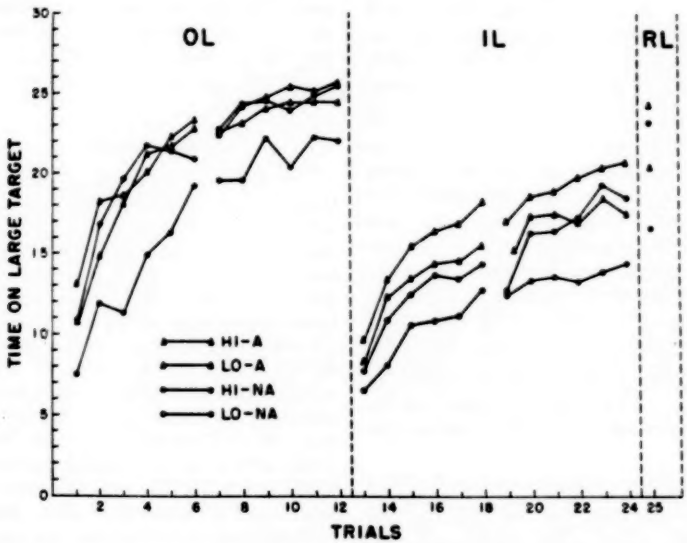

FIG 2. Performance curves for four groups of male *S*s during OL, IL, and RL practice on the Iowa Pursuitmeter, based on means of time on large target per trial.

that the LO-A and HI-NA group means are significantly different at the .5% level and the HI-A and HI-NA group means at only the 5% level. No other *t*'s approach significant values.

Since time on bullseye discriminated among *S*s during OL, and time on large target during IL, the OL bullseye scores and the IL large target scores were made comparable in order to obtain information regarding performance changes from OL to IL. The sum of the time on bullseye scores was obtained for each *S* over all OL trials and the distribution of scores was transformed into *Z* scores having a mean of 50 and *SD* of 10. A similar transformation was made of the large target scores for IL. It was found that the performance levels of the groups shifted relative positions from OL to IL practice. The HI-A group performed at a higher level on OL than on IL, the LO-A group higher on IL than on OL. A similar, but lesser shift occurred in the performance of the HI-NA and LO-NA groups, with the LO-NA group performing better relative to the HI-NA group during IL than during OL.

A Treatments × Levels analysis of the *Z* score data indicated that the observed interactions are not statistically dependable.

The traditional measure of retroactive effects is "percentage loss in performance." The data for percentage loss from the last OL trial to the RL trial are summarized in Table 1. The HI-NA group shows the largest loss, the LO-A group the smallest.

A simple analysis of variance of the percentage loss data based on time on large target was made. It indicated that the differences in group means, and by inference the differences in retroactive effects, are significant at the 5% level. Use of *t* tests indicated that the difference between the means of the LO-A and HI-NA groups is significant at the 1% level, and the LO-NA and HI-NA groups at the 2% level. No other *t*'s approach significance.

TABLE 1

TIME ON TARGET ON LAST OL AND FIRST RL TRIALS, AND PERCENTAGE LOSSES IN PERFORMANCE

Group	Time on Target (Sec.)		Loss
	Last OL Trial	First RL Trial	%
Bullseye target			
HI-NA	3.73	2.13	42.90
HI-A	6.49	4.17	35.75
LO-NA	4.58	3.33	27.29
LO-A	6.61	5.02	24.05
Large target			
HI-NA	21.99	16.61	24.47
HI-A	24.33	20.48	15.82
LO-NA	25.38	23.03	9.26
LO-A	25.54	24.29	4.89

DISCUSSION

The results indicate that among analyzers, those *S*s with a high drive level have larger proactive and retroactive losses on this task than do *S*s with a low drive level. Among non-analyzers, *S*s with a high drive level also show larger losses than do *S*s with a low drive level. These results can in large part be accounted for in terms of habitual modes of task approach, differences in difficulty between the OL and IL tasks, and the interaction of these factors with drive level. Studies in verbal learning (**4, 5, 12, 13**) indicate that drive level interacts with task difficulty in such a way as to facilitate performance on an easy task and impair performance on a difficult task. This interaction

has been interpreted by some as being due to the enhancing properties of drive. It is assumed, when the task is easy, that *S* has more facilitative response tendencies than response tendencies incompatible with proficient performance, and when the task is difficult, more incompatible than facilitative response tendencies. Drive energizes *all* response tendencies elicited in a situation, thus causing an interaction between task difficulty and drive level.

Since there is no practical way of determining the entire hierarchy of response tendencies possessed by the "average *S*," nor of determining all of the tendencies relevant to performing pursuitmeter tasks, it is not possible to predict the effect of differences in drive level on OL task performance. It is known, however, that the reversed (IL) task is much more difficult than the standard (OL) task. Therefore, high drive *S*s might be expected to do less well relative to low drive *S*s on the reversed task than on the standard task.

Similarly, it is not possible to predict the effects of drive on the OL performance levels of analyzers and non-analyzers. The reasons are: (*a*) the importance of an "analytical" approach for proficient pursuitmeter performance as compared with the importance of all other relevant tendencies in *S*'s response hierarchy is not known, and (*b*) it is not known whether the average *S* has more facilitative than incompatible response tendencies for the OL task.

It follows that high drive level could result in facilitation of performance for both analyzers and non-analyzers, impairment of performance for both groups, or facilitation of performance for analyzers and impairment of performance for non-analyzers. The data indicate that for early OL pursuitmeter practice, the last is the case. Since the reversed pursuitmeter task is more difficult than the standard task, it would be predicted that high drive analyzers would do less well relative to low drive analyzers, and high drive non-analyzers less well relative to low drive non-analyzers, on the reversed task than on the standard task.

These effects on performance arising from an interaction of drive and task difficulty would be expected if either task were practiced separately. When the tasks are practiced in sequence, additional effects would be expected to occur as a result of proaction and retroaction.

Since analyzers are assumed to gain more habit strength for responses related to underlying features of the task which are common to both the standard and reversed forms, they should show more positive transfer than non-analyzers do. It is known that the average *S* shows large proactive and retroactive losses in performance on the Pursuitmeter. It is assumed that this is due to his having learned more response tendencies during OL (proactive losses) or during IL (retroactive losses) which are not, than which are, facilitative of performance on both tasks. An analyzer learns more responses facilitative of performance on both tasks than the average *S*, and a non-analyzer fewer than the average *S*.

If drive energizes all response tendencies, these differences between the response strengths of analyzers and non-analyzers would be enhanced by a high drive level. Because proactive and retroactive losses of the average *S* are large on the Iowa Pursuitmeter, it seems likely that all groups would learn more response tendencies which are not than which are, facilitative of proficient performance on both tasks. If this is true, it would follow that high-drive *S*s would show more proactive and retroactive loss than low-drive *S*s. Low-drive analyzers would show the least loss, and high-drive non-analyzers the most loss. Analyzers would show less loss than non-analyzers.

Summary

In an attempt to explore the marked differences in transfer effects among individuals practicing complex motor tasks, 68 male *S*s, differing in achievement imagery scores and in

mode of approach to Block Design problems were given practice on the Iowa Pursuitmeter. All *Ss* had 12 standard task OL trials, 12 reversed task IL trials, and 1 standard task RL trial.

The *Ss* identified as "analyzers" from their verbal statements concerning their approach to block design problems performed at a higher level on the pursuitmeter during both OL and IL and showed less retroactive loss than "non-analyzers."

Low-drive *Ss* performed at a higher level than high-drive *Ss* except during early OL where high-drive analyzers performed more proficiently than low-drive analyzers. RL measures indicated differences in retroactive effects for the groups, high drive non-analyzers showing the largest, and low drive analyzers the smallest performance losses.

References

1. Barch, A. M., & Lewis, D. The effect of task difficulty and amount of practice on proactive transfer. *J. exp. Psychol.*, 1954, **48,** 134–142.
2. Cattell, R. B. The riddle of perseveration: I. "Creative effort" and disposition rigidity. *J. Pers.*, 1946, **14,** 229–238.
3. Darroch, J. Variation in the score of a motor perseveration test. *Brit. J. Psychol.*, 1938, **28,** 248–262.
4. Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52,** 311–327.
5. Hurley, J. R. Verbal learning as a function of instructions and achievement motivation. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.
6. Lewis, D., Adams, J. A., & Spieth, W. An analysis of performance on the Iowa Pursuit apparatus. Air Training Command Human Resources Research Center, *Res. Bull. 53-5*, March, 1953.
7. Lewis, D., & McAllister, D. E. An investigation of individual susceptibility to interference. U. S. Navy, Office of Naval Research, *Tech. Rep. SDC 938-1-10*, 1950.
8. Lewis, D., McAllister, D. E., & Bechtoldt, H. P. Correlational analysis of the learning and relearning of four different tasks on the modified Mashburn apparatus. *J. Psychol.*, 1953, **36,** 83–109.
9. Lewis, D., McAllister, D. E., & Bechtoldt, H. P. Correlational study of performance during successive phases of practice on the standard and reversed tasks on the SAM complex coordinator. *J. Psychol.*, 1953, **36,** 111–126.
10. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
11. Miles, G. H., & Lewis, D. Age and handedness as factors in the performance of a complex pursuit task: results of a study at the Iowa State Fair. *Proceed. Ia. Acad. Sci.*, 1956, **63,** 568–575.
12. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45,** 91–96.
13. Raymond, C. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46,** 120–124.
14. Spearman, C. *The abilities of man.* New York: Macmillan, 1932.
15. Spieth, W., & Lewis, D. The effects of alternating practice on the performance of two antagonistic motor tasks. U. S. Navy, Office of Naval Research, *Tech. Rep. SDC 938-1-6*, 1950.
16. Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.

(Received February 26, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

INFLUENCE OF TASK COMPLEXITY AND INSTRUCTIONS UPON SIMPLE AND DISCRIMINATION REACTION TIMES [1]

JOSEPH B. SIDOWSKI

San Diego State College

ROSS MORGAN, AND GORDON ECKSTRAND

Aero-Medical Laboratory, Wright Air Development Center

The present study was designed in order to investigate the influence of task complexity and instructions upon simple and discrimination RT.

Outside of the laboratory, there are few instances in which an individual is required to make a simple reaction. Generally, discrimination reactions are required. In addition, the tasks which must be performed after the initial reaction may vary greatly in complexity. An operator, for example, might be required to learn to associate the appearance of a light with the manipulation of a single control. However, the appearance of a second light might signify that the operator immediately manipulate a *number* of controls in a specified order. One would expect that an increase in the *number of alternatives* to which the operator could react would result in an increase in RT (**4, 7**). One might also expect that an increase in the *complexity of the task* (an increase in the number of controls to operate following the appearance of the light) would have some influence on response latency, since *S* might develop a different "preparatory set" for simple and complex tasks. A review of the literature, however, shows little experimental evidence to support the contention that complex movements produce longer RT's than do simple movements (**6, 7**).

METHOD

Subjects.—The *S*s were 48 right-handed male college students from the University of Dayton who were paid for their services. All *S*s had normal color vision.

Apparatus.—On a table, facing *S*, was an upright panel containing three colored stimulus lights (red, green, and amber), and three electrical switches (rotary, push-pull, and toggle). A reaction key was on the table directly in front of *S* within easy reach. A 6-v. AC buzzer was used as a signal, and a 2-sec. timer, using a relay operated thyratron circuit, served as a control for the duration of presentation for each of the stimulus lights.

Procedure.—A visual stimulus of constant duration was presented on each trial and *S* responded by lifting his hand off of a key.

A buzzer, activated by *E*, served as a ready signal. Each *S* responded to the buzzer by pressing down on the reaction key with the first two fingers of the right hand. After a randomly presented interval of 1.5, 2.0, or 2.5 sec., one of the stimulus lights was presented. The onset of the light acted as a signal for *S* to release the key. The light was presented for a 2-sec. interval.

After releasing the key, *S* made one of three possible responses which differed in task complexity. The simplest response was a simple finger withdrawal from the key. The second response was one in which *S* released the key and manipulated the center switch on the panel (a one-switch response). The third response was one in which *S* released the key and then operated all three switches on the panel.

The 48 *S*s were randomly assigned to two groups of 24 *S*s each. The groups differed only in the type of instructions which were presented. Group 1 was told that this was an experiment in RT and that *S* was to make his RT as short as possible during the course of the experiment.

[1] The data for this article were collected at the University of Dayton under Contract AF33 (616)-2763, and have been reported previously in Wright Air Development Center Technical Report 57-50.

Although *S* was also required to manipulate panel switches on two-thirds of the trials, he was told it was most important that he remove his fingers from the key as quickly as possible since *E* was primarily interested in recording RT measures. Reaction time was defined as the time interval between the onset of the stimulus light and the initiation of the response of removing the fingers from a reaction key. Group II was told to obtain as fast a total movement time as possible, and that *E* was primarily interested in recording the speed of total movement. Total movement time was defined as the time from the onset of the light to completion of the trial. It consisted of switch manipulation time as well as RT.

Each *S* in both groups served under a discrimination and a simple reaction condition. The *S*s served under one condition during the morning session and under a second condition during an afternoon session with the conditions being counterbalanced over sessions. During each session there were 90 trials which were divided into three blocks of 30 trials each. During the simple reaction condition, the center light *only* was used throughout 90 trials. No discrimination was required and *S* made the same response each time that the light flashed for 30 consecutive trials. For example, for 30 consecutive trials *S* made a simple finger withdrawal response each time that the light appeared. During the second block of 30 trials he manipulated the center switch on the panel, while during the last 30 trials *S* manipulated all three switches. All six permutations of the three task complexities were used with each *S* being assigned to one of the combinations.

During the discrimination condition, all three stimulus lights were used and *S* had to learn to associate a different response with each light. For example, if the red light flashed *S* was to make a simple finger withdrawal response and nothing more. When the green light appeared, however, *S* was to remove his hand from the key and manipulate the center switch. He was to react to the amber light by manipulating all three switches. The lights were presented randomly throughout each block of 30 trials over the 90 trials. The only limitation was that each of the three lights appeared an equal number of times in any block of trials. The presentation of blocks of trials and the arrangement of the colored lights from left to right on the control board were counterbalanced throughout the study.

A practice session of 12 trials and a 3-min. rest interval preceded each block of 30 trials.

Three time measures were recorded. Reaction times, in milliseconds, were measured by a Berkeley Electronic Counter, while movement times were measured in .01-sec. units by two Standard Electric timers.

Results

Reaction time measures.—The mean RT's are presented in Fig. 1. The curves indicate longer RT for the discrimination condition under both types of instructions. Reaction time instructions appear to produce shorter RT under all conditions. An analysis of variance for partially hierarchical models (**3**) was carried out on the data and is presented in Table 1. No significant Instruction effect was found but the difference between Reaction Types (Discrimination and Simple) was found to be significant at the .01 level. A significant difference was also found in Task Complexity indicating a difference in RT between the three possible response types. A significant interaction effect was found between Reaction Type and Task Complexity indicating that the Discrimination RT was more affected by response complexity than was the Simple RT. Duncan's test (**1**) for differences between ranked means was carried out to compare the three types of responses. Under Simple Reaction conditions, for both types of instruc-

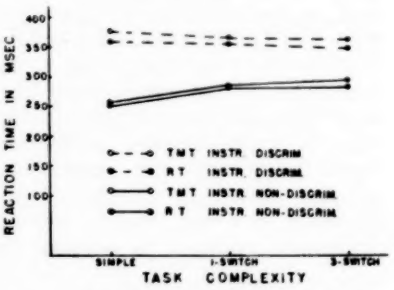

Fig. 1. Discrimination and non-discrimination RT in milliseconds as a function of task complexity under reaction-time (RT) and total-movement-time (TMT) instructions.

TABLE 1

ANALYSIS OF VARIANCE FOR REACTION TIME DATA

Source	*df*	*MS*	*F*
Reaction Type (RT)	1	464.60	147.21**
Instructions (I)	1	6.31	.57
Task Complexity (TC)	2	2.55	6.68**
Ss/Instructions (*Ss*/I)	46	11.16	
RT × I	1	1.45	.46
RT × TC	2	12.45	26.22**
RT × *Ss*/I	46	3.16	
I × TC	2	.05	.14
TC × *Ss*/I	92	.38	
RT × I × TC	2	.16	.35
RT × TC × *Ss*/I	92	.48	
Total	287		

** $P = .01$.

tions, RT's for the simple response were found to be significantly shorter than RT's for the one-switch response (.01 level) and for the three-switch response (.01 level). No significant difference was found between the one-switch and three-switch responses for the Simple Reaction condition. Under Discrimination conditions, the only significant difference in RT was found between the simple and three-switch response for Total Movement Time Instructions (.05 level), with the simple response being significantly longer than the three switch response.

In order to determine the effects of practice upon RT a comparison was made between the first block of five trials and the last block of five trials. An analysis of variance for repeated measures was carried out on these data. A significant Trial effect ($F = 50.42$, $df = 1$ and 46, $P = .01$) indicated that practice did have an effect upon RT while a significant difference between Reaction Types ($F = 130.13$, $df = 1$ and 48, $P = .01$) indicated that RT's were longer under Discrimination conditions. A significant interaction between Trials and Reaction Type ($F = 26.01$, $df = 1$ and 48, $P = .01$) showed that the Reaction Types were influenced differentially by practice with RT decreasing quite markedly over trials for the Discrimination conditions but not so rapidly for Simple RT.

A simple analysis of variance was carried out to determine if there was a difference in RT to the three colored lights. No significant difference between colors was found ($F = .04$, $df = 2$ and 94).

Movement time measures.—An analysis of variance for hierarchical models (**3**) was carried out on the movement time data. A significant difference was found between Reaction Types ($F = 106.01$, $df = 1$ and 46, $P = .01$) as well as between Task Complexity ($F = 1{,}448.15$, $df = 1$ and 46, $P = .01$). No other significant differences in movement time were found.

The mean movement times for the first block of five trials and the last block of five trials were subjected to an analysis of variance for repeated measures. The results were similar to those found for RT measures over trials. A significant Trial effect was found ($F = 93.47$, $df = 1$ and 46, $P = .01$); a significant difference between Reaction Types was found ($F = 82.77$, $df = 1$ and 46, $P = .01$); and a significant interaction between the above variables was found ($F = 31.63$, $df = 1$ and 46, $P = .01$). No other significant differences were found.

The results of the present experiment indicate that RT's differed significantly as a function of the reaction type (simple or discrimination) and as a function of task complexity. Reaction times did not differ significantly as a function of instructions.

Discussion

Since the significance of discrimination and simple reaction conditions and practice effects over trials are discussed elsewhere (**2, 7, 8**), the discussion will be limited to the task complexity variable.

On the basis of an earlier review of RT studies by Woodworth (**7**), a significant difference in RT as a function of task complexity would be expected. Woodworth's review suggested that complex movements produce longer RT's than do more simple moves, and that a movement aimed at a particular point is slower in starting than a full movement. A recent review by Teichner (**6**), however, states that available experimental evidence shows no apparent relationship between RT and consequent movement.

The present experiment shows that for simple reaction conditions, the simple response results in significantly faster RT's than either of the more complex switch movements. This would appear to support Woodworth's contention that complex movements and aimed movements produce longer RT. If, however, this contention is accepted as holding generally, there would be difficulty in explaining the results for the discrimination condition wherein the simple response for the total movement time instructions resulted in *longer* RT than the three switch response. One possible explanation for the longer discriminative RT's with simple responses may be found in the "preparatory set" developed by *S*. Under discrimination conditions, *S* had to be prepared to react to any one of three lights. It might be assumed that as a result of the reduced preparedness which the discrimination condition allowed for any one particular response, *S* directed his emphasis upon preparing for the most difficult response, the three-switch response. When, however, a simple or one-switch response was required, *S* had to clearly inhibit this established reaction tendency. The inhibition of the three-switch response, therefore, would result in a delay in reaction and hence a longer RT for the less complex responses.

Summary

The present experiment investigated the influence of task complexity and instructions upon simple and discrimination RT. The standard procedure for measuring RT was used; however, after releasing the key *S* made one of three possible responses which differed in complexity. The three responses ranged from a simple finger withdrawal to a one-switch manipulation response, to a more complex three-switch response. Group I ($N = 24$), with RT instructions, was told that this was a RT experiment and that *E* was primarily interested in measuring the fastest RT possible. Group II ($N = 24$), with total movement time instructions, was told to direct attention to obtaining as fast a total movement time as possible (total movement time included both RT and switch manipulation times). Each *S* in both groups served under simple reaction and discrimination RT conditions.

Analysis of the data showed no significant differences in RT between instructions. Discrimination conditions were found to produce longer RT's than did simple reaction conditions. Task complexity was also found to have a significant influence upon RT. The influence of task complexity was found to be a function of reaction type, discrimination or simple.

References

1. Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1042.
2. Farber, I. E., & Spence, K. W. Main and interactive effects of several variables on RT. *Technical Report 3, ONR*, Contract N9-onr-93802, 1955.
3. Harter, H. L., & Lum, M. D. Partially hierarchal models in the analysis of variance. *USAF, WADC Tech. Rep. 55-33*, 1955.
4. Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.
5. Poulton, E. C. Perceptual anticipation and reaction time. *J. exp. Psychol.*, 1956, **51**, 98–102.
6. Teichner, W. H. Recent studies of simple reaction time. *Psychol. Bull.*, 1954, **51**, 128–150.
7. Woodworth, R. S. *Experimental psychology* New York: Holt, 1938.
8. Woodworth, R. S. & Schlosberg, H. *Experimental psychology*. New York: Holt, 1954.

(Received March 11, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

EFFECTS OF TASK DIFFICULTY, STIMULUS SIMILARITY, AND TYPE OF RESPONSE ON STIMULUS PREDIFFERENTIATION[1]

BENNET B. MURDOCK, JR.

University of Vermont

This study investigated the effects of three variables, task difficulty, intra-list stimulus similarity, and type of response on stimulus predifferentiation (PD). As generally used PD refers to the positive transfer effect obtained in an A—B, A—C paradigm with visual stimuli and a low level of inter-list response interference. The facilitation presumably occurs because in some way *S* "learns about" the stimuli in Task 1 and this then aids performance in Task 2. Studies of PD have used both recognition and learning tasks as Task 2; the experiments reported here used only learning tasks.

Experiment I was conducted to determine if PD is a function of the difficulty of the task from which transfer occurs or of the task to which transfer occurs. Variations in difficulty were obtained by varying the presentation rate in Task 1 and in Task 2.

Experiment II was conducted to determine if PD is a function of intra-list stimulus similarity. A common explanation of PD is that, as suggested by Gibson (**2**), intra-list generalization tendencies are reduced in learning Task 1; this then would make Task 2 easier to learn. If reduction of intra-list generalization tendencies were a factor accounting for the positive transfer then the transfer should increase as intra-list stimulus similarity increases. This is so because as similarity increases the relative ease of learning responses to predifferentiated stimuli as opposed to non-predifferentiated stimuli should also increase.

Experiment III was conducted to determine if the transfer is a function of the type of response learned to the stimuli in Task 1. The Miller and Dollard concept of acquired distinctiveness of cues (**8**) has been suggested as a possible explanation of PD (**5**). Since the Task 1 response functions as a mediating response in Task 2, this interpretation would predict differences in transfer as a function of the type of response used.

General Procedure

All three experiments used the same basic condition, and this will be described first. Then, the specific variations within each experiment will be reported separately.

On Task 2 *S* had to learn a paired-associate task consisting of eight S—R pairs. The eight stimuli were patterns of lights in a 5×5 matrix of lights. There were eight push-buttons before *S*, and he had to learn which was the correct button to depress for each pattern. Prior to this, on Task 1 *S* had learned the "names" [100% association value nonsense syllables (**4**) FAM, GUL, NOV, WIX] for four of the eight patterns. The four stimuli used in Task 1 formed the stimuli for the four experimental (E) pairs in Task 2; those not used in Task 1 were the stimuli for the four control (C) pairs in Task 2.

Transfer was determined by comparing the learning of E and C pairs on Task 2. Each *S* served as his own control, and E and C pairs were reversed for half the *S*s. This design has been referred to elsewhere as Design IV with counterbalancing (**9**).

The lights were housed in a box approxi-

[1] This research was supported by a grant, G-2590, from the National Science Foundation. Mr. James Barnard tested the *S*s and helped in the analysis of the data.

mately 9 in. square, and the distance between each row and each column was 1.5 in. The box was approximately 12 in. to *S*'s left and level with the opening in a memory drum which was used to present the nonsense syllables. The *S* was seated, and memory drum and matrix were at eye level approximately 3 ft. away. The buttons were mounted on a response panel 15 in. wide and 9 in. deep. In all there were 12 buttons arranged from left to right in three groups of four each. Only the outer two groups were used here, and these two groups were mounted in semi-circles so *S* could comfortably keep the four fingers of each hand (the thumb was not used) on the buttons. All *S*s used the "touch" system rather than the "hunt and peck" system. The response panel was on a table between *S* and the memory drum.

For both tasks a standard paired-associate procedure was followed. On Task 1 a stimulus would appear in the matrix and then, a predetermined time later, the correct nonsense syllable would appear in the opening in the memory drum. On Task 2 a light directly behind the correct button would come on and stay on until the next stimulus appeared. The anticipation method was used on both tasks. In Task 2 the correct button had to be depressed at the time the light behind it came on. Then and only then a buzzer sounded signaling both to *S* and *E* the occurrence of a correct response.

In selecting the two sets of stimuli three criteria were used: (*a*) The intra-set similarity should be relatively high. This follows from Gibson's analysis; for reduction of generalization there rather obviously must be some generalization tendencies to start with. (*b*) The inter-set similarity should be low. To the extent that this criterion was not met there would be decreased transfer because of generalization between E and C pairs. (*c*) The two sets should be comparable in ease of learning, even though counterbalancing was used. One of the two sets actually used is shown in the first row (HS) of Fig. 1. The second set consisted of inverted V's; that is, the same patterns rotated 180°.

On List 2 the four buttons on *S*'s left were randomly paired with the upright patterns, the four buttons on *S*'s right with the inverted patterns. Also, on all trials the inverted patterns were presented in odd-numbered serial positions while the upright patterns were presented in even-numbered serial positions. These two features (using different hands for E and C pairs and alternating serial positions for E and C pairs) were employed to further reduce any generalization between E and C pairs. With counterbalancing, any systematic bias here should cancel out.

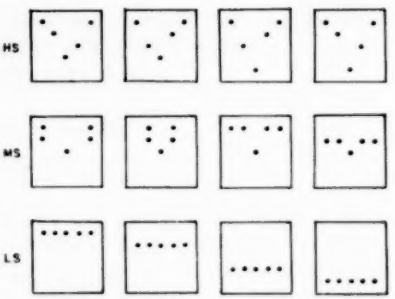

FIG. 1. High (HS), medium (MS), and low-similarity (LS) stimuli.

One trial consisted of the presentation of 10 S—R pairs. Since List 2 was 8 pairs long 2 pairs had to be presented twice in one trial; however, in every four trials each individual pair appeared five times. Since List 1 was 4 pairs long each individual pair appeared two or three times in one trial but five times in two trials. Four different presentation orders were used to discourage serial learning. The intertrial interval was 20 sec., and approximately 2 min. elapsed between the end of Task 1 and the start of Task 2.

On Task 1 all *S*s learned to a criterion of 19 out of 20 correct responses (i.e., two perfect trials with one mistake allowed). On Task 2 all *S*s learned to a criterion of one perfect trial or for 20 trials, whichever came first. The *S*s for all three experiments were drawn from the same population, students of both sexes enrolled in the introductory psychology course, and were randomly assigned to groups.

EXPERIMENT I

Method.—In Task 1 there were two presentation rates, 1.5 sec. (i.e., 1.5 sec. for S, 1.5 sec. for S—R) and 4 sec. In Task 2 there were four presentation rates, 1.5 sec., 2 sec., 3 sec., and 4 sec. Each *S* learned Task 1 at one of two rates and Task 2 at one of four rates. With this 2×4 design there were 6 *S*s per cell, (Total $N = 48$), and within each cell counterbalancing was used.

Results.—The measure of transfer was given by the formula $T = E - C/E+C$ (**9**), where E and C are the total number of correct anticipations of the E and C pairs on Task 2. The mean values are shown in Table 1. An analysis of

TABLE 1

PERCENTAGE TRANSFER

Task 2 Rate	Task 1 Rate				
	1.5 Sec.		4 Sec.		Combined Mean
	Mean	*SD*	Mean	*SD*	
1.5 sec.	33.3	9.1	35.2	12.7	34.2
2.0 sec.	20.5	18.9	18.2	11.4	19.3
3.0 sec.	2.5	15.6	29.5	10.2	16.0
4.0 sec.	22.3	10.3	5.5	14.5	13.9

variance[2] showed that the Task 2 rate did have a significant effect on transfer ($F = 4.85$, $P < .01$). Tukey's gap test (**1**) showed that the 1.5-sec. rate gave significantly ($P < .05$) more transfer than the other three rates, but the three slower rates did not differ among themselves. The effect on transfer of the Task 1 rate was not significant ($F < 1.00$), but the interaction was ($F = 4.78$, $P < .01$). The interaction seemed to occur largely at the two slower Task 2 rates; whether a 3-sec. or a 4-sec. rate gave more transfer depended on which Task 1 rate had been used.

If the data for the two Task 1 groups are combined, each of the four Task 2 groups showed significant positive transfer (t's of 10.21, 4.09, 2.81, and 3.06 for the 1.5-, 2-, 3-, and 4-sec. groups, respectively; for all cases $P < .05$). In the 2-, 3-, and 4-sec. groups 18 *S*s reached criterion in less than 20 trials and 18 did not; the difference between them in transfer was insignificant ($t < 1.00$).

An analysis of Task 1 learning showed that the 1.5-sec. group required over twice as many trials to reach criterion as the 4-sec. group, and this difference was highly significant ($F = 28.30$, $P < .01$). However, the four Task 2 groups did not differ among themselves in Task 1 learning, which suggests that the groups were similar in learning ability. Also, almost exactly the same number of trials were required to learn the nonsense syllables to the upright patterns as to the inverted patterns, thus indicating that the two sets of patterns were of comparable difficulty.

[2] The justification for using parametric statistics on this measure of transfer is given elsewhere (**10**).

EXPERIMENT II

Method.—In this experiment three different levels of intra-list stimulus similarity were used, and for each level one of the two sets is shown in Fig. 1. To obtain the other set the high similarity (HS) and medium similarity (MS) stimuli were rotated 180°, while the low similarity (LS) stimuli were rotated 90°.

Two criteria were used in determining level of similarity: (*a*) number of identical elements; and (*b*) redundancy. The HS, MS, and LS sets of stimuli had two (first and last light), one (middle light), and no identical elements, respectively. The HS, MS, and LS stimuli were 40%, 60%, and 80% redundant. That is, in the HS set two of the five lights (first and last) were unnecessary for proper identification of the particular pattern. In the MS set three were unnecessary; the middle light and either the left or right hand two (which were always symmetrical). In the LS set four of the five were unnecessary; any one light was sufficient to identify the pattern.

In part, these two criteria were chosen because, ". . . *indistinguishability*, rather than *similarity* or *homogeneity* . . . is the important variable" (**3**, p. 95). It seems reasonable that indistinguishability is correlated with the number of identical elements. Greater redundancy should give *S* more ways of identifying the stimuli and so make it easier to discriminate them, i.e., reduce indistinguishability.

A 1.5-sec. presentation rate was used in both Task 1 and Task 2. All *S*s were given 20 trials on Task 2 regardless of performance. A total of 42 *S*s was employed.

Results.—The group means were 18.0%, 3.4%, and 8.8% for the HS, MS, and LS conditions, respectively. An analysis of variance showed that the similarity variable did not have a significant effect on transfer ($F = 1.84$, $P > .05$). Of the three

groups only HS showed significant positive transfer ($t = 4.18$, $P < .01$).

In terms of number of trials to learn Task 1 to criterion the group means were 22.1, 10.0, and 6.1 trials for HS, MS, and LS, respectively. As the means and variances were correlated a logarithmic transformation was used; an analysis of variance of the transformed scores showed that these differences were significant ($F = 27.57$, $P < .01$). By Tukey's gap test both the HS—MS and the MS—LS differences were significant ($P < .01$, one tail). Thus, in so far as Task 1 learning is an indication of intra-list stimulus similarity, it would certainly appear that the three groups of stimuli really did differ in similarity.

Experiment III

Method.—To determine if transfer is a function of the type of response, the Task 1 responses were either verbal or motor. To control for inter-task relationships the Task 2 responses were also either verbal or motor. With this 2 × 2 design, two of the four groups (verbal-verbal and motor-motor) would have incompatible responses in the two tasks and inter-list interference should be high. The other two groups (verbal-motor and motor-verbal) would have compatible responses and interference should be low.

The verbal Task 1 has already been described. The motor Task 1 used the four push buttons mounted in the middle third of the response panel already described. The buttons were positioned so the first two fingers of each hand could rest on them comfortably; again the "touch" system was used. The lights for the middle set of buttons had red jewels while those for the two outside sets of buttons had green jewels.

The motor Task 2 has already been described. The verbal Task 2 used eight new 100% association value syllables (KEN, REG, TIG, VAC, DIF, HOB, JAN, PED).

For the motor-motor group the pairing of stimuli and response was done so that in no case was either the same finger on the same button correct for a given stimulus in both tasks. The presentation rate was 1.5 sec. throughout, the HS stimuli were used, and all Ss learned Task 2 for 20 trials. There were 12 Ss in each of four groups.

Results.—The mean transfer values were: verbal-motor, 24%; verbal-verbal, 20%; motor-verbal, 0%; motor-motor, 4%. An analysis of variance showed that the type of Task 1 response did affect transfer ($F = 5.29$, $P < .05$). However, neither the type of Task 2 response nor the interaction was significant (for both $F < 1.00$).

Both the verbal-motor and the verbal-verbal groups showed positive transfer (t's of 2.81 and 2.76; for both, $P < .05$). However, neither the motor-verbal nor the motor-motor group differed significantly from zero (for both, $t < 1.00$). The results clearly indicate that, irrespective of response compatibility, a Task 1 verbal response gives positive transfer while a Task 1 motor response does not.[3]

On Task 1 learning, the verbal group required a mean of 23.8 trials to reach criterion while the motor group required a mean of 33.8 trials. This difference was significant ($F = 7.95$, $P < .01$). However, the Task 2 variable was not significant, suggesting equality of learning ability.

Discussion

Experiment I showed that PD is a function of the difficulty of Task 2, and the most difficult condition (1.5-sec. rate) gave the most transfer. These results seem quite reasonable. At one extreme Task 2 could be so easy that any prior training would be superfluous, and at the other extreme Task 2 could

[3] The zero transfer found for the Task 1 motor groups is supported by the results of a previous unpublished study. With a 3-sec. rate, different stimuli (symmetrical and asymmetrical V's), and the Star-D as the motor response, both a motor-motor ($N = 10$) and a motor-verbal ($N = 10$) group showed essentially zero transfer (2.8% and 1.4%, respectively).

be so difficult that it could not be mastered without prior training. Of course, presentation rate is only one way of varying task difficulty.

Experiment II failed to confirm a prediction, based on Gibson's theory, that PD would increase with intra-list stimulus similarity. Differences in similarity that were great enough to have a clear-cut effect on (Task 1) learning were not great enough to have a significant effect on transfer. However, the significant transfer found for the HS group directly contradicts Goss's prediction (5) that stimuli high in similarity should show little positive (or perhaps even negative) transfer.

Experiment III showed that PD is a function of the type of Task 1 response, and verbal responses result in more transfer than motor responses. One explanation for these results is that, in learning Task 2, it would have been easier to use the previously learned verbal responses than the motor responses. However, this explanation seems unlikely; the availability of these mediating responses in learning Task 2 would presumably depend in part on the degree of response compatibility, yet compatibility was unrelated to transfer. Also, *S*s in the motor-verbal group could certainly have made use of implicit Task 1 responses in learning Task 2, yet this group showed no sign of positive transfer.

In an experiment on mediated generalization Jeffrey (7) assumed that response-produced verbal cues were more distinctive than response-produced motor cues, and found more mediation with verbal responses. This suggests a different explanation for the results of Exp. III; there was more transfer with the verbal responses because their response-produced cues were more distinctive.[4] In connection with this it is relevant to note that the verbal responses on Task 1 were easier to learn than the motor responses. This would be expected if the verbal cues were more distinctive than the motor cues.

Finally, the results of Exp. III suggest that while reduction of intra-list generalization tendencies may be a necessary condition for PD it is not a sufficient condition for PD. In Task 1 the motor group learned responses to the same stimuli and to the same criterion as the verbal group; therefore the intra-list generalization tendencies at the start of Task 2 learning should have been comparable for the two groups. Furthermore, the stimuli were high in similarity (note the number of trials required to learn Task 1 to criterion) so reduction of generalization tendencies should be a critical factor. Yet despite this the verbal group showed considerable positive transfer while the motor group showed none.

Summary

Three experiments were conducted to determine the effects of task difficulty, stimulus similarity, and type of response on stimulus predifferentiation. Experiment I showed that transfer increased as the Task 2 presentation rate increased; Exp. II failed to find an increase in transfer as intra-list stimulus similarity increased; Exp. III showed that verbal responses learned on Task 1 gave more transfer than motor responses, but that inter-list response compatibility did not affect transfer. It was suggested that the distinctiveness of the response-produced cues may be a significant variable, and that reduction of intra-list generalization tendencies, while perhaps a necessary condition, is not a sufficient condition for stimulus predifferentiation.

References

1. Edwards, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.
2. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

[4] Hake and Eriksen (6) found that the number of mediating responses available did not affect transfer, and this clearly speaks against the importance of response similarity. Their arguments notwithstanding, it is possible that in their experiment differences among groups in degree of Task 1 learning canceled out the effects of differences in distinctiveness due to number of mediating responses available.

3. Gibson, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28,** 93–115.
4. Glaze, J. A. The association value of non-sense syllables. *J. genet. Psychol.*, 1928, **35,** 255–269.
5. Goss, A. E. A stimulus-response analysis of the interaction of cue-producing and instrumental responses. *Psychol. Rev.*, 1955, **62,** 20–31.
6. Hake, H. W., & Eriksen, C. W. Effect of number of permissible response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50,** 161–167.
7. Jeffrey, W. E. The effects of verbal and nonverbal responses in mediating an instrumental act. *J. exp. Psychol.*, 1953, **45,** 327–333.
8. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
9. Murdock, B. B., Jr. Transfer designs and formulas. *Psychol. Bull.*, 1957, **54,** 313–326.
10. Murdock, B. B., Jr. "Backward" associations in transfer and learning. *J. exp. Psychol.*, 1958, **55,** 111–114.

(Received March 13, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

MOTOR CONFLICT BEHAVIOR AS A FUNCTION OF MOTIVATION AND AMOUNT OF TRAINING[1]

BURTON G. ANDREAS[2]

State University of Iowa

A number of experiments using human *Ss* have been based on the assumption that motor conflict arises from competition between tendencies to make physically incompatible motor responses (**1, 3, 7, 10**). Different cues have been used separately in training two antagonistic motor responses; then the simultaneous presentation of the cues has been considered to generate conflict by arousing tendencies to make both responses at once. Blocking, hesitation, and reversals of direction of overt responses are among the response characteristics which indicate conflict.

The assumption that the amount of conflict is a function of the absolute strength of equally strong competing tendencies in human *Ss* has been tested several times, not always with success. Sears and Hovland (**10**) increased the strengths of the competing tendencies by using shock avoidance as incentive in one experiment, and by giving additional training trials in another, but in neither case was the assumption confirmed. (The report of these two experiments does not indicate that this assumption was being tested explicitly, but the data given bear directly upon it.) Bilodeau (**3**) did find conflict to be a function of the amount of pre-conflict training when she used movement time, or number of incomplete responses plus number of blockings as two different measures of conflict.

Sears and Hovland (**10**) explicitly tested the assumption that conflict decreased as the competing tendencies became more unlike in strength. This was confirmed when shock was used to manipulate motivation, but only partly confirmed when amount of training was the independent variable. In the latter experiment a group given 20 trials on one response and one trial on the other showed significantly less blocking than groups trained equally on the two responses. However, a 20-to-5 ratio of training trials did not yield significantly less blocking than the equal training groups. Bilodeau (**3**) found support for the assumption when she used unequal numbers of training trials on the two responses (to yield unequal competing tendencies).

It would appear that two points of procedure utilized in the experiments mentioned might profitably be altered. The conflict trial involved the simultaneous presentation of two stimulus lights which had previously been experienced singly. This, of course, has always been a conventional step in testing motor conflict, but it has the disadvantage of giving *Ss* a situation which has not been covered by instructions. This may elicit extraneous responses which complicate the competition of the two motor response tendencies. Instructing *Ss* to perform the two competing responses may also have been an undesirable procedural feature. This may have raised the avoidance tendencies so close to the limit of their growth that the subsequent practice trials did little to affect the strengths of the competing response tendencies. This might account for some of the difficulty in confirming basic conflict hypotheses. Because of these possible difficulties, it appears desirable to test conflict theory in a situation in which (*a*) the simultaneous pres-

[1] This article is based on a dissertation submitted to the faculty of the Department of Psychology of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to thank Professor I. E. Farber for guidance in conducting the research. This study was reported in part at the 1954 APA meeting.

[2] Now at the University of Rochester.

entation of cues is not unexpected by *S* although motor conflict is still generated, and (*b*) avoidance tendencies are gradually built up through trial-and-error learning rather than by instructions. The proposed procedure resembles one employed by Spence (**12**) in studying discrimination learning in chimpanzees.

The aim of the present experiment was to assess the conflict generated by a manipulation of (*a*) motivation and (*b*) amount of training using a procedure modified according to the above suggestions. Two assumptions were tested: (*a*) with strengths of competing response tendencies approximately equal, conflict increases as the strengths increase in absolute value; (*b*) conflict increases as the strengths of unequal tendencies approach equality.

Method

Subjects.—The *S*s were 264 students enrolled in introductory psychology. Men and women were assigned in equal numbers to 11 groups of 24 *S*s each.

Apparatus.—The apparatus modeled after one described by Siipola (**11**) has been described elsewhere (**2**). A handle could be moved freely from a starting point to any one of 13 target positions. The target positions were notches spaced around a semicircular framework which rested on a table. Stimuli were green pilot lights located at the target positions. Seven of these stimuli and their corresponding target positions were used in this experiment. They will be referred to by numbers, 1 to 7, assigned consecutively beginning at *S*'s left. These numbers did not appear on the apparatus and were not mentioned in the instructions. A yellow warning light located beyond the center stimulus light (No. 4) was presented .2 sec. prior to any stimulus light. A green light indicated to *S* when the handle was in starting position. Movement of a microswitch at the starting position opened a Standard Electric timer circuit permitting measurement of response latency from the onset of stimuli to the initiation of the overt motor response. A second timer measured movement time from starting point to any target position. Screens concealed the control and recording apparatus from *S*. A buzzer sounded for 1.5 sec. to signal the making of a correct response during training. A shock circuit delivered a shock of 1.5-sec. duration to the left wrist when an error was made.

Procedure.—The *S*s were given seven familiarization trials on which they were instructed to move the handle as quickly as possible from the starting point to the target position indicated by the stimulus light which flashed on, the seven lights being presented once each in irregular order. New instructions then informed *S* that two stimulus lights would flash on together on subsequent trials and that he was to move as quickly as possible to one of them. Electrodes were placed on his left wrist and the level of shock was tested and adjusted so that *S* reported a mild pain sensation when shock was administered. All *S*s used the right hand to move the handle. A "correct" response resulted in the sounding of the signal buzzer, while a "wrong" response led to a shock.

On the training trials the stimulus lights were presented in pairs, with Light 2 or Light 6 a member of each pair. The target designated by either of these was always wrong. Targets indicated by Lights 1, 3, 5 and 7 were always correct. Each of these four was paired an equal number of times with each of the negative lights (2 or 6).

Eight groups of *S*s were given training series which involved Lights 2 and 6 equally often. The four series used contained 8, 16, 24, or 40 training trials prior to the first conflict trial. At each amount of training two groups were run, one at a moderate level (M), one at a high level (H), of motivation. The moderate level was attained by maintaining the shock just above *S*'s pain threshold throughout the experiment. The high level of motivation was reached by stating to *S*, four trials prior to the first conflict trial, that the intensity of the shock was being doubled. At this point the voltage was actually increased by 10 v.

On the first conflict trial, Lights 2 and 6 were presented simultaneously for the first time. This trial was undistinguished from prior training trials except for the particular lights used. A response to either target was punished at the increased shock level so as to maintain avoidance tendencies set up in training.

An additional training series, equal in length to the first, followed the first conflict trial without a break in procedure. For the high motivation groups another increase of 10 v. in the shock was introduced four trials prior to the second conflict trial. The *S* was told that the shock was now being turned up to maximum intensity.

The second conflict trial was a repetition of

the first, shock being given for a response to either Light 2 or Light 6. A third conflict trial followed immediately, but this one involved two positive stimuli, Lights 1 and 7.

The equal training groups, whose treatment has been described above, were given the following designations: 4:4M, 8:8M, 12:12M, 20:20M, 4:4H, 8:8H, 12:12H, and 20:20H. In each designation the ratio represents the number of training trials involving each negative stimulus prior to the first conflict trial. Thus, Group 8:8H was one which had eight trials involving Light 2 and eight involving Light 6 prior to the first conflict trial which was administered under high motivation. Additional treatment, described above, then followed.

Three additional groups of *Ss* were given unequal training under moderate motivation with one of the negative stimuli presented more often than the other. The ratios of training are indicated in the group designations: 4:20, 4:12, and 8:12. The preponderant use of Light 2 or 6 was counterbalanced within each group. Except for the unequal training, the pattern of treatment was the same as in the equal training groups run under moderate motivation: training series, conflict trial, additional training in the same ratio, two final conflict trials.

Results

Equal training groups.—For the conflict trials the frequency of occurrence of different types of responses indicating conflict was noted. Assuming that the number of reversals, partial responses, and blocks is a measure of conflict, the data suggested that conflict was greater in the groups tested under high motivation. Combining data from all three conflict trials, there was a total of 30 conflict-indicating responses in the high-motivation groups, compared with 21 made by the groups under moderate motivation. When data for both levels of motivation and all conflict trials were combined, it was found that conflict indices increased in frequency from 3 to 23 as level of training increased from the 4:4 groups to the 20:20 groups. The low frequencies of conflict-indicating responses plus the interdependence of

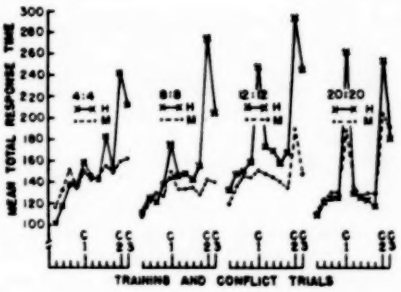

Fig. 1. Mean total response time on training and conflict trials for equal training groups. Values on the ordinate are hundredths of a second.

responses on successive conflict trials precluded the demonstration of statistical significance for the trends noted in the frequency data.

The mean total response time (reaction latency plus movement time) on training trials and on conflict trials is shown in Fig. 1 for the four equal training groups. C_1, C_2, and C_3 indicate the successive conflict trials.

The figure shows that the increase in total response time on conflict trials, when compared with training trials, is greater for the high motivation group than for the comparable training group performing under moderate motivation. A comparison of the curves for the four levels of training shows that the gain in response time on conflict trials tended to be an increasing function of the amount of preceding training.

Since the distribution of total response times was skewed, the reciprocals of these measures were taken to permit analysis of variance. As a reciprocal of movement time the new measure can be regarded as the speed of response; the gains in movement times on conflict trials now become decrements in response speed.

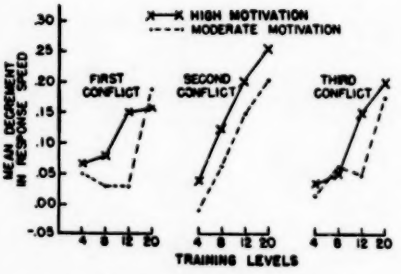

FIG. 2. Mean decrements in speed of response (1/T) on conflict trials for equal training groups.

Using the response speeds for three preceding training trials as a base, the mean decrements on conflict trials were calculated. They are represented in Fig. 2. The data for the first conflict trial were subjected to a 2×4 factorial-type analysis of variance. For the training variable, $F = 2.54$, $df = 3$ and 187, and P is slightly greater than .05. The motivation effect was clearly not significant for the first conflict trial.

An analysis of variance of the mean decrements in speed of response on the second conflict trial yielded $F = 2.97$, $df = 1$ and 187, and $P > .10$ for the motivation variable. The training effect proved to be very significant at this second conflict trial ($F = 8.46$, $df = 3$ and 187, $P < .001$).

Performance on the third conflict trial showed less decrement in speed of response than had been observed on the second conflict trial. Analysis of variance showed the effect of motivation to be not significant. The training variable was again significant ($F = 5.80$, $df = 3$ and 187, $P < .001$).

Unequal training groups.—The distribution of types of conflict responses made by the unequal training groups on the conflict trials was substantially the same as for the equal training groups. Single, direct approaches occurred predominantly, with a few reversals and no partial responses or blocks. Groups 4:12M and 8:12M made 9 reversals each, summing across the three conflict trials, while Group 4:4M made no conflict-indicating responses. Group 4:12M made 9 reversals, Group 8:12M made 3, and Group 12:12M made 6. These frequency data can be used neither to support nor deny the hypothesis that conflict is greater when competing response tendencies are closer to equality.

Decrements in speed of response on conflict trials were noted in the unequal training groups. Mean values are presented in Table 1, together with data from selected equal training groups for convenient comparison.

The data for Groups 4:20M, 4:12M, and 4:4M were analyzed together since these form a series varying in degree of inequality of competing tendencies, Group 4:4M being a minimal limit. For the first conflict trial there was no significant difference among the means of these three groups. For the second

TABLE 1

MEAN DECREMENTS IN RESPONSE SPEED (1/T) FOR UNEQUAL TRAINING GROUPS, WITH DATA FOR SELECTED EQUAL TRAINING GROUPS

($N = 24$ in each group)

Group	Conflict Trial		
	1	2	3
4:4M	.054	−.011	.021
4:12M	.032	.093	.078
4:20M	.055	.153	.150
4:12M	.032	.093	.078
8:12M	−.014	.067	.082
12:12M	.033	.148	.048

and third conflict trials, analysis of variance yielded F ratios of 3.88 and 4.86, respectively, indicating significant differences among the means ($.05 < P < .01$, $df = 2$ and 69). Since the mean decrement in speed of response was greatest on these two trials for Group 4:20M, these findings are inconsistent with the hypothesis that conflict increases as unequal competing tendencies approach equality.

For each of the three conflict trials, the performances of Groups 4:12M, 8:12M, and 12:12M were analyzed together. There were no significant differences in response speed decrements among these groups on any of the conflict trials.

Discussion

As compared with previous motor conflict studies (**1, 3, 7, 10**) the present experiment showed far less blocking and other responses indicative of conflict. This may be attributable to the simple fact that the simultaneous presentation of two lights was not a novel situation as had been the case in earlier work. Going further, it may be suggested that blocking in previous experiments did not arise from competition between motor response tendencies but stemmed from tendencies to make other responses. Hovland and Sears (**7**) themselves noted that verbal responses sometimes did actually replace manual responses on conflict trials.

On the other hand, the present experiment does confirm the finding by Bilodeau (**3**) that temporal response measures may relate conflict behavior to the strengths of competing tendencies. The usefulness of speed measures has been demonstrated in animal studies of conflict (**4, 6, 8**). These measures support the suggestion by Miller (**9**) that hesitation in responding is symptomatic of conflict.

With respect to the basic conflict assumptions that were tested, the present study found considerable support for the first, i.e., that with competing tendencies about equal, conflict is an increasing function of their absolute strengths. Both number of training trials and shock intensity were found to have some effect in raising these strengths. The fact that shock levels did not differentiate conflict performances to a statistically significant extent may be simply due to the fact that even the moderate shock was noxious enough to develop strong avoidance tendencies.

The second assumption about conflict, that it increases as competing tendencies approach equality, was not confirmed in this study. There was, in fact, some evidence against it, notably the appearance in Group 4:20M of the greatest decrement in speed of response when compared with Groups 4:4M and 4:12M on the second and third conflict trials. The assumptions tested in this study were essentially identical with those cited for frustration or conflict by Brown and Farber (**4**). They indicated (p. 484) that the importance of absolute vs. relative strengths of competing response tendencies was difficult to assess. The results of the present study suggest that greater importance attaches to the absolute values.

Summary

The assumptions that motor conflict is a function of the absolute and the relative strengths of competing response tendencies were tested. Competing tendencies were brought to different strengths by manipulating the amount of training and the motivational level of the college students who served as *Ss*.

The finding of far fewer instances of blocking than in earlier studies was attributed to the use of paired stimuli in training as well as on conflict trials.

When measured in terms of decrement in speed of response, conflict did appear to be greater for higher values of competing tendencies, especially when these were built up through training. No evidence was found for the assumption that conflict increases as unequal competing tendencies approach equality. The evidence was, rather, that the absolute level of avoidance tendencies is the primary determinant of conflict.

References

1. Andreas, B. G. Motor responses as a function of type of conflict. *Proc. Iowa Acad. Sci.*, 1950, **57**, 361–365.
2. Andreas, B. G. Empirical gradients of generalization in a perceptual-motor task. *J. exp. Psychol.*, 1954, **48**, 119–122.
3. Bilodeau, I. McD. Conflict behavior as a function of the strengths of competing response tendencies. Unpublished doctor's dissertation, State Univer. of Iowa, 1950.
4. Brown, J. S. Factors determining conflict reactions in difficult discriminations. *J. exp. Psychol.*, 1942, **31**, 272–292.
5. Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.
6. Bugelski, R., & Miller, N. E. A spatial gradient in the strength of avoidance responses. *J. exp. Psychol.*, 1938, **23**, 494–505.
7. Hovland, C. I., & Sears, R. R. Experiments on motor conflict. I. Types of conflict and their modes of resolution. *J. exp. Psychol.*, 1938, **23**, 477–493.
8. Kaufman, E. L., & Miller, N. E. Effect of number of reinforcements on strength of approach in an approach-avoidance conflict. *J. comp. physiol. Psychol.*, 1949, **42**, 65–74.
9. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
10. Sears, R. R., & Hovland, C. I. Experiments on motor conflict. II. Determination of mode of resolution by comparative strengths of conflicting responses. *J. exp. Psychol.*, 1941, **28**, 280–286.
11. Siipola, E. M. Implicit or partial reversion-errors: a technique of measurement and its relation to other measures of transfer. *J. exp. Psychol.*, 1940, **26**, 53–73.
12. Spence, K. W. Analysis of the formation of visual discrimination habits in chimpanzee. *J. comp. Psychol.*, 1937, **23**, 77–100.

(Received March 15, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

PERCEPTUAL EFFICIENCY AS RELATED TO INDUCED MUSCULAR EFFORT AND MANIFEST ANXIETY [1]

MILTON F. SHORE

Boston University [2]

Only recently has the experimental work on the effect of motivation on performance considered the relationship between drive level and effort (task difficulty and/or complexity). Most of the early work on the relationship between effort and performance has been summarized by Meyer (**6**). One of the ways of systematically inducing changes in effort is by changes in muscle tension. Meyer found that in almost all the studies using induced muscular tension, performance was facilitated, reached an optimum, then declined. The optimum point was a function of the type of muscular effort condition induced, and the nature and difficulty of the task.

Most of the studies on the effect of induced muscular effort on performance have dealt with learning tasks. Despite the increase in research on the effects of drive level on perceptual performance (see **8**), Shaw (**9**) has made the only report of a study of the effect of induced muscular effort on perceptual performance. As was found in previous studies on learning, he found that the perception span for digits increased then decreased as muscular tension was continuously increased. He also found that the optimal point was related to the task difficulty, the longer series requiring greater tension to reach optimal efficiency than the shorter series.

The purpose of the present study was to investigate the effect of effort (induced muscularly) and drive level (manifest anxiety) on visual recognition. To avoid difficulties inherent in the use of either affective materials or verbal stimuli, the stimuli used were standard photographic resolution targets. While the task remained constant, the effort conditions were varied systematically by induction of muscular tension. Differences in drive level were determined by the Taylor Manifest Anxiety Scale (MAS), a scale adapted from the MMPI and believed to be a measure of drive level (**10**).

METHOD

Subjects.—The Taylor Manifest Anxiety Scale (MAS) in its abbreviated form (the 50 anxiety items without the 175 filler items) was administered to over 300 undergraduate students (male and female) in psychology courses at Boston University. The abbreviated form has been shown by Mandler (**5**) to have correlations of .95 and .85 with the 225-item form in two independent studies.

From their scores on the MAS, three groups of six males were selected for the experiment: those with scores of 1–5 constituted the "Low Anxiety" (LA) group, those with scores of 13–15 the "Medium Anxiety" (MA) group, and those with scores of 24–28 the "High Anxiety" (HA) group. All *Ss* had 20/20 visual acuity corrected or uncorrected and all were

[1] This paper is based on a dissertation submitted to the Graduate School, Boston University, in partial fulfillment of the requirements for the Ph.D. degree. It was conducted in the Psychological Section, Physical Research Laboratory under contract with the Photographic Reconnaissance Laboratory, Wright Air Development Center, United States Air Force (Contract No. AF 33(616)-432). The author wishes to express gratitude to Dr. Leo J. Reyna for his direction and invaluable assistance.

[2] Presently at the Worcester Youth Guidance Center, Worcester, Massachusetts.

right-handed. The *Ss* were selected on the basis of time available for the experiment. An attempt was made to control the motivation induced by the incentive to participate in the experiment by telling *Ss* on the day of the practice session that they would receive a fixed sum of money for their time.

Apparatus.—The stimulus targets were presented in an electronically-timed tachistoscope adjustable for cumulative exposure speeds ranging from .001 to 10 sec. These targets were presented on $2\frac{1}{4} \times 2\frac{1}{4}$ in. opaque slides in a Kodachrome slide changer attached to a mask behind the tachistoscope. The circular opening in the center of the mask had a diameter of $1\frac{1}{2}$ in.

In this tachistoscope the pre-exposure lamps remained on when the exposure lamps flashed. Pre-exposure field illumination was 10.2 ft.-candles; exposure field illumination was 12.1 ft.-candles. Four lines perpendicular to the center of the sides of the field terminating 2 in. from each other in the center of the field were used for fixation.

A Smedley hand dynamometer was used to induce a graded series of muscular effort conditions. The dynamometer was modified by placing it on a mount and adding an electrode connected to the tachistoscope timer. In this way, when the pressure indicator touched the electrode, the tachistoscope would flash.

Stimulus materials.—Five standard photographic resolution targets drawn according to the specifications listed by Perrin and Altman (**7**) were used as stimuli for recognition. These were three parallel lines, divided square, Landolt **C**, annulus, and the Snellen **E**. The parallel lines and divided square were presented in both vertical and horizontal positions. The Landolt **C** was presented with the opening either to the left, right, top, or bottom. Thus there were 10 experimental targets. The *Ss* were told the target names used in the experiment during the practice session and before each experimental session. In order to reduce possible response dependencies and for use as measures of reliability, three targets were used which were not mentioned to *Ss*: a square without a dividing line, an asterisk, and a blank. All targets were $\frac{1}{4}$ in. along a side or in diameter, and the visual angle subtended was 17.9′.

Reliability.—To test inter-day variability, two *Ss* were presented the targets on four consecutive days under two tension conditions. Analysis of their scores showed no significant difference between the days.

Reliability of the daily performance of *S* was analyzed by the responses to the blank target (*Ss* were not told a blank would be presented). Of all responses to the blank target only .7% were an experimental target. Most of the responses were "nothing," "blur," or "blob." Therefore, *Ss* reliably reported what they saw.

Practice session.—In order to familiarize *Ss* with the apparatus and to reduce any variability caused by differences in learning ability, the first session was a practice session.

Following a test for visual acuity, *S* was seated at the tachistoscope and adjustments were made so that *S*'s optical axis was in the center of the aperture. Adjustments were also made in the dynamometer for a position most convenient for *S*'s hand size and arm length.

The *S* was then told that the purpose of the experiment was to determine some of the factors that influence visual perception. The operation of the tachistoscope was described and the 10 experimental targets were presented at a 1-sec. exposure speed. The *S* was corrected if he gave a name different from the one used in the study. The target names used in the experiment were "vertical square," "horizontal square," "vertical lines," "horizontal lines," "E," "doughnut," "C left," "C right," "C top," and "C bottom." The *S* was instructed to describe what he saw after each exposure even if it was only a flash or a blob. In order to encourage this he was told that targets might be presented that he had not seen before. Targets were then presented at exposure speeds of .001, .005, and .007 sec. The *S* was then told that following a response he was to say whether he was "sure," "doubtful," or "guessing." Three targets were then presented at .003, .007, and .017 sec. Following this, *S* was instructed in the use of the dynamometer and presented with targets at .005, .01, and .1 sec.

Following the practice trials, *S* was instructed to squeeze the dynamometer as hard as he could. A reading at the end of 15 sec. was taken as the maximum pressure for the day.

Experimental session.—Each of the six experimental sessions began with a maximum reading of dynamometer pressure at the end of 15 sec. Although pressure readings were taken each day, the maximum pressures for the six days varied only slightly, so that the pressure at a higher fractional level was at no time lower than the pressure at a lower fractional level. The *S* was then told: "Let's run through the targets again. (*E* presented the 10 targets at a 1-sec. exposure speed and corrected *S* if he had forgotten the target names.) Remember, there is a vertical square, a horizontal square, vertical lines, horizontal lines, an E, a doughnut, and a C with the opening to the right, left, top, and bottom. The targets will now be presented in random order. Each time the light flashes you are to tell me what you see. If it is a flash,

say so. If it is a blob, say so. If it is a target, tell me which one it looks like. Remember, there will be targets presented that you have not seen before. Each time you name something you have seen you are to state whether you are 'Sure,' 'Doubtful,' or 'Guessing.' When I say 'Press' you are to press the microswitch (or dynamometer). Try it." (*S* was given two examples at .003 and .007 sec.)

The targets were presented in an order randomized for exposure time and target. Inter-target presentation time was approximately 10 sec. Eighty experimental targets and 12 confusion targets were presented each day. Twenty-three targets (20 experimental targets, the blank, asterisk, and square) were presented at one sitting. Following each sitting, *S* had a 5-min. rest period to reduce fatigue and boredom. The daily sessions consisted of four sittings of 23 stimuli, a total of 92 presentations per day.

Each target was presented daily at .002, .004, .006, .008, .010, .012, .014, and .016 sec. To prevent recognition of the confusion targets with a high degree of certainty, the confusion targets were presented only at .002, .004, .006, and .008 sec. each day. The *S* was asked to draw what he saw when his description was not clear.

The fractional dynamometer pressures used were 0, $\frac{1}{8}$, $\frac{1}{4}$, $\frac{3}{8}$, $\frac{1}{2}$, and $\frac{3}{4}$ of the maximum pressure for the day. The order of presentation of the six fractional pressures was counterbalanced for each anxiety group.

The *S*'s score for the day was the number of targets correctly recognized. The data used in the analysis, therefore, consisted of the scores of six daily sessions taken from three groups of *S*s with six *S*s per group.

Results

The overall effect of induced muscular effort on the recognition of the targets is seen in the pooled curve in Fig. 1. Analysis of variance (Table 1) shows an *F* ratio for the effort conditions of 3.39, a value significant beyond the .01 level. Therefore, as in the studies on the effect of induced muscular effort on the efficiency of learning, as reported by Meyer (6), an increase in induced effort facilitates perceptual activity until an optimum is reached. Beyond this optimum, an increase in induced effort results in a progressive drop in efficiency.

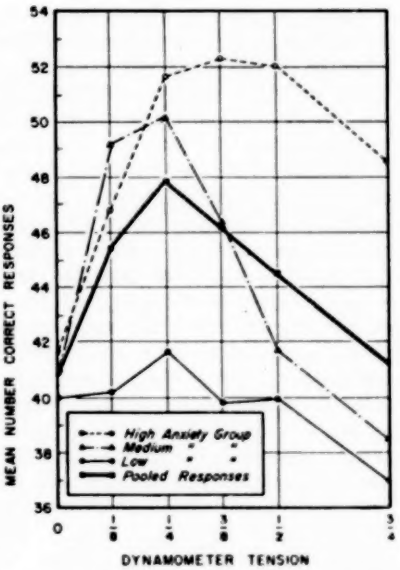

FIG. 1. Performance of the three anxiety groups on all targets under different conditions of induced effort.

The three anxiety groups, however, responded quite differently to the increase in induced muscular effort. The performance of the three groups is seen in Fig. 1. Analysis of variance (Table 1) shows an *F* ratio for the interaction of 3.12, a value significant beyond the .01 level. Induced muscular effort, therefore, interacted with manifest anxiety to produce dif-

TABLE 1

ANALYSIS OF VARIANCE OF RECOGNITION SCORES

Source	*df*	*MS*	*F*
Groups (G)	2	742.7	1.58
Conditions (C)	5	138.7	3.39**
G × C	10	127.6	3.12**
Ss within group	15	469.0	
Ss × C (pooled)	75	40.9	
Total	107		

** $P < .01$.

ferences in perceptual efficiency; Ss with high MAS scores showed an increase in efficiency of recognition, those with low MAS scores showed a slight decrease in efficiency as effort was increased.

The difference in the performance of the three anxiety groups is heightened by an evaluation of their responses to different target shapes. An equal number of circular and square targets were presented to S in this experiment. Square targets were found to be consistently more difficult to perceive than the circular targets. Comparison of the scores on the circular targets for the three groups revealed no significant difference between the groups. However, on square targets, as seen in Fig. 2, the differences between the groups are more marked than in Fig. 1. Analysis of variance (Table 2) shows an F ratio for the interaction of 2.77, a value significant beyond the .01 level. Thus target difficulty appears to summate with effort induced muscularly to interact with manifest anxiety influencing perceptual efficiency.

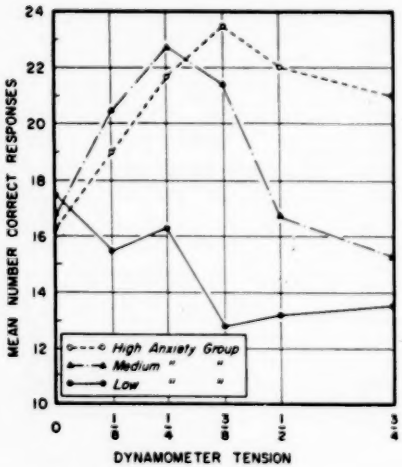

FIG. 2. Performance of the three anxiety groups on square targets under different conditions of induced effort.

TABLE 2

ANALYSIS OF VARIANCE OF RECOGNITION SCORES ON SQUARE TARGETS

Source	*df*	*MS*	*F*
Groups (G)	2	319.2	1.88
Conditions (C)	5	38.0	2.65*
G × C	10	39.6	2.77**
Ss within group	15	169.4	
Ss × C (pooled)	75	14.3	
Total	107		

* $P < .05$.
** $P < .01$.

DISCUSSION

The pooled results were consistent not only with the results of Shaw (**9**) on the perception of digits, and with the studies of the effect of induced muscular effort on learning reviewed by Meyer (**6**), but also resembled the results found in recent studies on intersensory stimulation (**11**). In the present experiment, it could be said that as proprioceptive stimulation and tactual stimulation were increased, the threshold for visual stimuli was first lowered then raised.

Deductions from these theories of intersensory stimulation, however, were inadequate to explain the differences in the performance of the three anxiety groups. These theories would hypothesize that the differences in the three groups resulted from differences in the tension induced, so that the LA group showed the inhibitory curve expected under high proprioceptive and tactual stimulation while the HA group showed the facilitation curve expected when proprioceptive and tactual stimulation are low. But a comparison of the maximum pressures for the three groups revealed almost identical mean pressures under all conditions. Therefore, the differences between the groups did not result from differences in stimulation.

Another hypothesis is that the groups differed in perceptual skill. If this were true, the differences should have been

present when the microswitch was used. Though there were slight differences between the groups at this point, these differences were far from significant.

A third hypothesis is that the increase in correct responses was a function of an increase in the number of guesses the *S*s were willing to venture at different levels of tension. Counting the responses of "flash" and "blob" as neutral and determining the number of positive errors for the three groups at the different tension levels produced results which were far from significant. The number of correct responses, therefore, could not be explained as a function of guesses the *S*s were willing to venture.

An explanation more consistent with the results obtained here is suggested by the recent work of Eriksen and Davids (**3**), Deese, Lazarus, and Hamilton (**1**), and Deese, Lazarus, and Keenan (**2**). These authors found a high correlation between the MAS and the Hysteria-Psychasthenia Scale of the MMPI. The results suggested that, rather than measuring drive level, the MAS may have cue properties which arouse anxiety in *S* towards which he responds in a manner characteristic of the way he handles stress situations. From their studies, one could hypothesize that *S*s low on the MAS would react to increased stress (induced muscular effort plus stimulus difficulty) by blocking out the environment and in this way raising their perceptual thresholds, while *S*s with high scores on the MAS, since they are highly sensitized to anxiety and anxiety-inducing stimuli, would show reduced perceptual threshholds. A hypothesis of this kind would be supported by the results of the present study.

Summary

This experiment was designed to test the effect of drive level and effort on perceptual efficiency. Non-affective stimuli were presented for recognition under six induced muscular effort conditions to three groups varying in drive level as measured by the Taylor Manifest Anxiety Scale.

The results revealed a high interaction between the two variables. Subjects with high MAS scores showed a significant increase in efficiency, while *S*s with low MAS scores showed a decrease in efficiency as induced effort was increased. The *S*s with median MAS scores showed a rise in efficiency to an optimal level, then a gradual decline. These results were more consistent with a theory based on the consistency of an individual's reaction to stress than with either theories of intersensory stimulation, theories that suggested differences in perceptual skill between the three groups, or theories that suggested an increase in the general response level of the organism.

References

1. Deese, J., Lazarus, R. S., & Hamilton, R. Anxiety and stress in learning: the role of intra-serial duplication. *J. exp. Psychol.*, 1954, **47**, 111–114.
2. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety reduction and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.
3. Eriksen, C. W., & Davids, A. The meaning and clinical validity of the Taylor anxiety scale and the Hysteria-Psychasthenia Scale of the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.
4. Lucas, J. D. The interactive effects of anxiety, failure and intra-serial duplication. *Amer. J. Psychol.*, 1952, **56**, 59–66.
5. Mandler, G. Personal communication.
6. Meyer, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.
7. Perrin, F. H., & Altman, J. H. Studies in the resolving power of photographic emulsions: VI. The effect of the type of pattern and the luminance ratio in the test object. *J. opt. Soc. Amer.*, 1953, **43**, 780–790.
8. Postman, L. The experimental analysis of motivational factors in perception, in *Current theory and research in motivation: a symposium.* Lincoln: Univer. of Nebraska Press, 1953.
9. Shaw, W. A. Facilitating effects of induced tension upon the perception span for digits. *J. exp. Psychol.*, 1956, **51**, 113–117.
10. Taylor, J. A personality test for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
11. Trehub, A. A theory of sensory interaction: an experimental investigation of the relationship between autonomic activity and visual sensitivity. Unpublished doctor's dissertation, Boston Univer., 1954.

(Received March 18, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

EVIDENCE THAT THE SECONDARY REINFORCING STIMULUS MUST BE DISCRIMINATED

F. J. McGUIGAN AND FRANCES CROCKETT[1]

Hollins College

According to Keller and Schoenfeld (**6**), a stimulus must first be established as a discriminative stimulus in order to act as a secondary reinforcer for any response. The importance of this hypothesis is apparent when one notes that this condition is not usually included as a necessary condition for the establishment of a secondary reinforcer. Hull, for instance, states that a neutral stimulus needs only to occur ". . . repeatedly and consistently in close conjunction with a reinforcing state of affairs . . . [in order to] acquire the power of acting as a reinforcing agent" (**5,** p. 6). If in fact the hypothesis advanced by Keller and Schoenfeld is true, it would seem that the prior establishment of a stimulus as a discriminative stimulus is sufficiently important to be included as a necessary condition in such general statements as Hull's. The evidence for the discriminative stimulus-secondary reinforcement hypothesis is, however, somewhat limited (cf., **3, 6, 8, 10, 11**). The argument in most of these studies seems to proceed somewhat as follows: (*a*) a stimulus (which is not established as a discriminative stimulus) is repeatedly associated with primary reinforcement; (*b*) it is found that that stimulus does not acquire secondary reinforcing powers; (*c*) it is therefore concluded that a necessary condition for a stimulus to become a secondary reinforcer is that it first be established as a discriminative stimulus. However, a more satisfactory basis for reaching this conclusion would be as follows: For one group of *S*s, the stimulus is established as a discriminative stimulus (at which time the stimulus is associated with reinforcement), while for a second group the stimulus is not established as a discriminative stimulus, but is merely associated with reinforcement. If it is then found that the stimulus becomes a secondary reinforcer for the first group, but not for the second group, there would be support for the conclusion that the prior establishment of the stimulus as a discriminative stimulus is a necessary condition for its serving as a secondary reinforcer. The present experiment was designed to test the discriminative stimulus-secondary reinforcement hypothesis following this latter procedure.

METHOD

Subjects.—The *S*s were 48 male and female albino rats from the Sprague-Dawley Strain, approximately 90–120 days old.

Apparatus.—The training apparatus was adapted from that reported by Grice (**4**). The *S* went from a starting box through an alley from which he could continue forward and enter one of two end boxes. There were four possible end boxes, any two of which could be used depending on the experimental condition, by sliding the starting box and entrance alley at right angles to the direction of the alley. Three of the end boxes were black and one was white for half of the experiment, while for the other half three were white and one was black. Guillotine-type doors were placed at the exit end of the starting box, and at the entrance ends of the end boxes.

[1] The authors wish to acknowledge the assistance of Carolyn Bolton, Eleanor Boylan, and Llewellyn White.

The testing apparatus was a Y maze with removable end boxes. One end box was white and the other black.

In both the training and the testing apparatus the alleys and starting doors were a neutral gray, and the end box doors were the same color as the end boxes.

Procedure.—All *Ss* were placed on a daily diet of approximately 8 gm. of Purina Laboratory Chow about one week prior to the training sessions. They were fed once a day at about the time they were to have their maze runs. The *Ss* were then randomly assigned to two experimental conditions, with the exception that an equal number of males and females were assigned to each condition. Group I was trained under a discriminative stimulus condition while Group II was trained under a non-discriminative stimulus condition. In the training maze, the *Ss* of Group I ran to either a black or a white end box. For half of these *Ss* the white box contained food, while the black box did not; for the other half the black box contained food, while the white box did not. In Group II, half of the *Ss* ran to one of two black boxes, while the other half ran to one of two white boxes; in either case both boxes contained food. The color of the box that contained food will be referred to as the "positive stimulus" and the color of the box that did not contain food will be referred to as the "negative stimulus." All *Ss* were detained for 20 sec. in the end box to which they ran; those who ran to the positive stimulus were allowed to eat for that period of time. All *Ss* ran 5 trials a day for five days, and 10 trials a day for six days. The positive stimulus appeared for *Ss* in Group I an equal number of times on each side in a random order. The *Ss* in Group II were forced in a backward serial order to the opposite side of their first five choices for the last five trials of the last six days, i.e., if the first five trials were RRLRL, they were forced RLRLL on the last five trials.

A necessary condition was that the *Ss* in Group I not be reinforced in the presence of the negative stimulus. As a result, if they did not run to the negative stimulus for several trials, they were forced to do so in a random manner. This, of course, was no problem when black was negative, but was when white was negative.

Immediately after completing the day's trials, all *Ss* were removed to a distinct cage for 20 min., and then returned to their home cages where they were fed.

On Day 12 all *Ss* were run in the Y maze for 10 trials. The number of times that they ran to the white and the black end boxes was recorded. The end boxes were placed randomly to the right and left sides an equal number of times, neither of which ever contained food.

Results

Two criteria of secondary reinforcement in Y-maze performance on Day 12 were used: the frequency with which *Ss* went to the positive stimulus on Trial 1, and the frequency with which each *S* went to the positive stimulus on the first 10 trials. Thus, if *S* went to the positive stimulus six or more times during the 10 trials he was classified as showing a preference for the positive stimulus; otherwise *S* was classified as not showing a preference for the positive stimulus.

All statistical tests were made with the use of the binomial distribution. There were no significant test performance differences between *Ss* that had black as the positive stimulus during training and the corresponding *Ss* that had white as the positive stimulus. The results were thus combined and are presented in Table 1. Using the first trial as the criterion of secondary reinforcement it was found that neither Group I nor Group II went significantly more often to the positive stimulus than to the negative stimulus. Furthermore, there was no significant difference between the two groups. According to the first criterion, then, it may be concluded that neither group showed evidence of secondary reinforcement. However, using the second criterion, frequency of choices during the first

TABLE 1

Frequency of Choices of the Positive Stimulus on Criterion Trials

Group	*N*	First Trial		All 10 Trials	
		S^+	S^-	S^+	S^-
I	24	16	8	22	2
II	24	13	11	14	10

10 trials, it was found that Group I went to the positive stimulus significantly more often than chance ($P < .01$). Group II, however, did not go to the positive stimulus significantly more often than chance. A comparison of the two groups indicated that Group I made significantly more choices of the positive stimulus than did Group II ($P < .05$). It may therefore be concluded, according to the second criterion, that the positive stimulus acquired the power of acting as a secondary reinforcing stimulus for Group I, but not for Group II.

Discussion

According to the discriminative stimulus—secondary reinforcement hypothesis, the positive stimulus should have acquired the capacity of acting as a secondary reinforcing stimulus for Group I, but not for Group II. This is so because the positive stimulus was established as a discriminative stimulus for Group I, but not for Group II. Since it was found, according to the second criterion, that the positive stimulus became a secondary reinforcer for Group I, but not for Group II, it may be concluded that the discriminative stimulus-secondary reinforcement hypothesis was confirmed. It may be noted that the results obtained by Saltzman (**9**) although not directly related to the Keller-Schoenfeld hypothesis are in accord with the present findings.

Two important problems directly relevant to the present study are the definition of a "discriminative stimulus," and a specification of the methods for its development. The difficulty in providing a unique definition of the discriminative function of a stimulus has been considered elsewhere (**1, 2, 3, 6, 7, 8**). As to the development of a discriminative stimulus, one suggestion is that it ". . . involves not only the reinforcement of responses made in the presence of a positive stimulus, but the withholding of reinforcement for responses made in the absence of the positive stimulus or in the presence of a negative stimulus" (**10,** p. 44). It is well to ask, however, what is meant by ". . . withholding of reinforcement for responses made in the absence of the positive stimulus . . .," for practically all of the time that *S* is not being reinforced in the presence of the positive stimulus, reinforcement is withheld in the absence of the positive stimulus. For example, according to a strict interpretation of this procedure, the positive stimulus is being established as a discriminative stimulus for our "non-discriminative stimulus group" (Group II), *since practically all the time* S *is not actually in the positive end box, reinforcement is withheld and the positive stimulus is absent,* as, for instance, occurs when *S* is in his home cage. One way to help clarify this matter is to suggest that the withholding of the reinforcement must occur in the absence of the positive stimulus *in the experimental situation.* Thus, *in a given situation,* while *S* is in the presence of the positive stimulus, he is reinforced, and when *S* is not in the presence of the positive stimulus, he is not reinforced. It follows, then, that one characteristic of a discriminative stimulus is that it is one to which a response is generally made to one stimulus in a given situation, but the response is generally not made to any other stimulus in that situation. It should be emphasized that while these considerations may help, they do not solve the problem, for many additional questions present themselves, e.g., how can we more precisely define "a given situation"? (one that is limited spatially and temporally?); what are the relative characteristics of the two stimuli that are differentially reinforced, i.e., must they be on some sort of common continuum (such as color, black vs. white), or can they be totally different, such as color of the end box vs. size of the end box?

These results and considerations suggest, although they are certainly not conclusive, that a necessary condition for the establishment of a stimulus

as a secondary reinforcer is that the stimulus be first established as a discriminative stimulus.[2] As a result, it is suggested that general principles of secondary reinforcement, such as that of Hull's, be modified to read somewhat as follows: In a given situation, limited both spatially and temporally, if one neutral stimulus is repeatedly and consistently associated with a reinforcement, whether primary or secondary, and a different neutral stimulus is not similarly associated with a reinforcement, then the first stimulus will acquire the power or acting as a reinforcement.[3]

Summary

The hypothesis that a stimulus must first be established as a discriminative stimulus before it can become a secondary reinforcing stimulus was tested. Two groups of rats were trained such that a stimulus was established as a discriminative stimulus for Group I, but not for Group II. To accomplish this Group I ran to either a black or a white end box, only one of which contained food. Group II ran to one of two end boxes of the same color, both of which contained food. Performance in a Y maze which had a black and a white end box, neither of which contained food, was used as the test of secondary reinforcement. Two criteria were used: frequency of choice on the first trial and on the first 10 trials. Performance on the first trial was not an effective criterion, but the second criterion indicated that the stimulus that had been associated with reward became a secondary reinforcer for Group I, but not for Group II. It was thus concluded that the prior establishment of a stimulus as a discriminative stimulus is a necessary condition for that stimulus to become a secondary reinforcer. The relevance of this conclusion to a general principle of secondary reinforcement was discussed, as was the problem of defining a discriminative stimulus and specifying the methods for establishing a discriminative stimulus.

[2] There are a number of experiments in which it is not readily apparent that discriminative training on the neutral stimulus was given, yet that stimulus became a secondary reinforcer. A discussion of the relationship of such experiments to this conclusion is beyond the scope of the present paper. The interested reader might refer to the discussion of this matter by Schoenfeld, Antonitis, and Bersh (**10**), who point out that such experiments cannot be unequivocally interpreted. It is also possible that a neutral stimulus which has not been established as a discriminative stimulus may still acquire secondary reinforcing properties, but that our procedures for testing for secondary reinforcers are not sufficiently refined to yield positive results. The results of Saltzman (**9**) are somewhat suggestive that this may be the case.

[3] This modification is offered with respect to the problem dealt with in this experiment only, and is thus minimal in nature. A complete statement of a principle of secondary reinforcement would require the consideration of a large number of problems beyond the scope of the present paper (cf., **7**).

References

1. Bugelski, B. R. *The psychology of learning.* New York: Holt, 1956.
2. Dinsmoor, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *J. exp. Psychol.*, 1950, **40**, 458–472.
3. Ferster, C. B. The effect on extinction responding of stimuli continuously present during conditioning. *J. exp. Psychol.*, 1951, **42**, 443–449.
4. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination. *J. exp. Psychol.* 1948, **38**, 1–16.
5. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
6. Keller, F. S., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.
7. McGuigan, F. J. The logical status of Hull's principle of secondary reinforcement. *Psychol. Rev.*, 1956, **63**, 303–308.
8. Notterman, J. M. A study of some relations among aperiodic reinforcement, discrimination training, and secondary reinforcement, *J. exp. Psychol.*, 1951, **41**, 161–169.
9. Saltzman, I. J. Maze learning in the absence of primary reinforcements: a study of secondary reinforcement. *J. comp. physiol. Psychol.*, 1949, **42**, 161–173.
10. Schoenfeld, W. N., Antonitis, J. J., & Bersh, P. J. A preliminary study of training conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1950, **40**, 40–45.
11. Webb, W. B., & Nolan, C. Y. Cues for discrimination as secondary reinforcing agents: a confirmation. *J. comp. physiol. Psychol.*, 1953, **46**, 180–181.

(Received March 18, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

MUSCLE RESPONSES AND THEIR RELATION TO ROTE LEARNING[1]

R. N. BERRY AND R. C. DAVIS

Indiana University

Level of muscle activity and its changes come into question in a number of interpretations of learning. A high level of activity, it has been proposed, makes learning more rapid, either because it improves connection forming or merely because it lowers response threshold (**4**). It has also been suggested (**1**) that muscular activity *per se* has no univocal effect of this sort. *Changes* in level, following signals informing *S* of the correct response, are of interest in connection with well-known learning theories. In a "reinforcement" interpretation, "tension reduction" is expected to occur after a response has been signalled to be correct. Although skeletal muscle activity is not the only possible referent of the general term "tension," it is a possible meaning. By contiguity analysis, one would expect some "correction" response to occur when *S* receives a signal that he should have made a different response. Since by one theory activity should decrease when *S* receives a confirmation and since by the other it should increase (or decrease less) after he receives a corrective signal, there is little to separate the two predictions. The overt activity of making a response will tend to produce a higher muscular activity temporarily, which would be expected to decline to a greater or lesser degree thereafter, depending on what kind of a signal *S* receives.

It is the purpose of the present experiment to ascertain some relations between muscular activity and learning, relations which may be expected by a number of theories, rather than to search for critical differences between theories. The principle relations dealt with are: (*a*) between muscular activity and successive time units within a syllable presentation; (*b*) between activity level and the learning scores of individuals; (*c*) between change of level following a response and the learning scores; and (*d*) between differential effects following right and wrong responses and learning scores.

Although the effect of an induced muscular tension upon learning has been studied in a number of experiments (**4**), there is a scarcity of information about the relation of normally occuring tension, which naturally may be quite different. In this experiment, therefore, data were obtained by recording muscle action potentials from three surface locations while *S*s performed a standard serial learning task.

METHOD

Subjects.—The *S*s were 24 male undergraduate students, volunteers from the elementary psychology course. The only specific effort to supply motivation was a statement that cooperation with the experimenter would result in extra course credits.

Procedure.—Each *S* was placed in a chair in the experimental room; the room contained a memory drum and was separated from a recording room by a sound-reducing partition. Electrodes with electrode paste were fastened

[1] This study was carried out under Contract Nonr-908(03). Dr. R. C. Davis is the Principal Investigator under this contract. The senior author is indebted to the Graduate School of Indiana University for financial assistance in carrying out this study.

on the forehead (frontalis), right forearm (extensors), and jaw (masseter) muscles of *S*. These were connected to amplifiers in the recording room. This recording equipment, previously described by Ellson, Davis, Burke, and Saltzman (2), provides a permanent polygraph record of the muscle action potential, integrated over .1-sec. intervals.

The *S* was seated so that his eyes were 3 ft. from the aperture of the memory drum. The aperture was illuminated evenly by means of light sources invisible to *S*. The memory drum mechanism was set to move the drum every 3 sec. On the drum there were nine nonsense syllables, selected from Glaze's (3) zero-association list. Each syllable was presented for 3 sec. with a 3-sec. interval of blank white paper between syllables. Following the last syllable of the list were two blanks giving a total of 6 sec. between trials of the list. A red line was put in the second of these blanks, preceding the first syllable in the list; the *S* was instructed to respond with the first syllable, by spelling it aloud, when this red line appeared, and thus to anticipate each syllable during the blank interval before its appearance. Emphasis was put on the point that no vocal responses were to be made while a syllable was in the aperture. The *E* told *S* that there would be 20 presentations of the list no matter how well or how poorly he performed but that he was to try to learn the list as fast as possible. If *S* inquired as to the significance of the electrodes, he was simply informed that *E* was interested in the action of the muscles during learning. All *S*s were reassured that there would be no shock administered at any time.

The list of syllables (TOV, CEF, GAX, YEM, VAF, SIJ, MEF, ZIW, QAP) was changed from *S* to *S* in such a way that every syllable fell in each of the nine possible serial positions. At least two *S*s were presented with each of the serial orders.

Scoring.—The polygraph tracings were divided into 3-sec. intervals corresponding to the intervals on the memory drum. All 3-sec. intervals were further divided into two equal segments of 1.5 sec. The maximum MAP was then measured in log microvolt units (×100) for each of these segments for each of the three electrode placements. There were, then, four segments of each 6-sec. unit. The first two segments represent the vocal "response interval" and the last two the "information interval."

Right responses were defined as those in which *S* spelled the three letters of the syllable, in their correct order, during the 3 sec. response interval. A wrong response was defined as one in which *S* vocalized any other set of letters than the correct ones; if no understandable response was made, the response was characterized as neither right nor wrong. In order that the action potentials associated with right and wrong responses might be compared, readings for the first 10 right and the first 10 wrong responses were selected, this number of responses being the total available for some *S*s[2]. These readings form the principal body of data for the study.

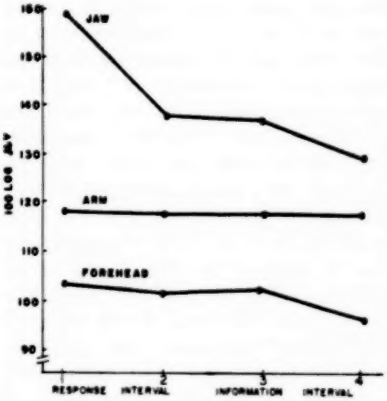

FIG. 1. The mean MAP values for each electrode placement for the four segments of the 6-sec. interval spanning the vocal response and information periods.

RESULTS

Temporal relations.—The MAP means for the four segments of the 6-sec. interval, averaging values of right and wrong responses, are shown in Fig. 1. For the jaw and forehead, the values are lower for the information intervals. For the arm there is a remarkable identity of all means in spite of considerable variation among the individual readings. Differences in absolute level of activity among the several locations are much influenced by local factors, but temporal

[2] Two *S*s were discarded from the original 26 because one of them had only three incorrect responses and the other only six incorrect responses. It is recognized that the exclusion of these *S*s may bias the data; however, their inclusion would create a bias of a different sort.

changes, with data in log form, may well be compared. Evidently there is a difference between response and information intervals in the head region, particularly in the focal site of the response. This is the expected outcome for these locations.

Levels of muscular activity and learning.—The general level of muscular activity for *S* for a single location is defined as the mean of all readings in the body of data analyzed. In the principal data, therefore, right and wrong responses, and response and information intervals, are equally represented. In the comparison of *S*s on the basis of such averages, one would expect a considerable variance to arise from electrode placement, body structure, and other irrelevancies. But inasmuch as these may be taken as random with respect to such a factor as learning achievement, they produce no bias but do make significance hard to establish.

The relation between the level of activity in the several locations was examined by computing rank difference correlations among individuals. Jaw and forehead locations correlate to the extent of .54 ($P < .01$), whereas the correlations of the arm with the other locations are near zero and non-significant. As jaw and forehead seem to have a good deal in common, readings for the two, for each *S*, were therefore averaged to simplify the treatment and increase reliability. Arm readings were treated separately.

In Fig. 2 general level readings, jaw and forehead combined, are broken into four components, right and wrong responses, and response and information intervals, and each component is plotted as a function of learning score. Learning scores were defined as the total number of correct responses made in the 20 learning

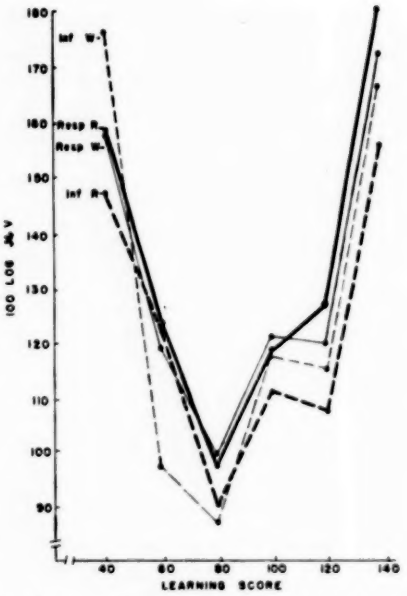

FIG. 2. The MAP values (jaw and forehead combined) for response and information intervals subdivided into right and wrong categories are related to learning scores.

trials and were classified into 20-point intervals. The mean action potential of all *S*s in an interval is plotted at the midpoint of the interval. (In the 6 intervals, from low to high, the number of *S*s is 3, 2, 5, 6, 6, 2.) All 4 sets of readings treated in this fashion form **U**-shaped curves, with the best and the poorest learners having high action potentials. To test the significance of this trend it is, of course, necessary to use a statistic which does not assume a linear relation. The non-parametric Kruskal-Wallis test (**5**, pp. 185–193) was chosen. This should indicate whether the scores in the six learning categories would reasonably be represented by a horizontal line through the general median. Values tested by this technique were the general

activity level measures, i.e., all four classes of readings were averaged for each *S*. It is inconvenient to get an exact *P* value for the resulting statistics because the small numbers in certain categories prevent the direct application of the chi-square distribution. The table provided by Kruskal and Wallis (**5**, pp. 282–283), however, shows that with small category numbers the chi-square distribution is too broad, i.e., *P* values read from it are too large. In the present case the chi-square value is .07. Since the probability is too large, and since the test in any case takes no account of orderly sequence, it is reasonable to reject the hypothesis that the readings in the six learning groups are drawn from the same population. There is, therefore, in this sort of situation some relation between activity level and learning performance. Further, the relation is not monotonic, since the rank-difference correlation between the variables is only .16, far from any standard of significance.

Mean trial position of the readings differs for the several groups of *S*s, i.e., good learners produced 10 right responses sooner than the poorer *S*s. A progressive adaptation effect could, therefore, have had some effect on the curves of Fig. 2, although it could not produce an inversion of the trend. A special set of readings was prepared in which responses were so selected as to give the same mean trial position for the readings for each *S*. (Naturally this resulted in fewer readings for each *S*.) In the plot of these data the **U**-shaped curve is as prominent as before, though, as would be expected, because of reduced number of readings, it is less significant.

Readings from the first 10 "no response" presentations (where the *S* gave no understandable response) further confirm the shape of this function. Readings of this sort, principally from the first showing of each syllable, are related in the same **U**-fashion to the learning scores subsequently earned by *S*s. The "no response" readings tend to be somewhat lower than the others, particularly for the extreme groups. The data are not presented because of space limitations and because the definition of "no response" is such that the individual could have made either a very large *ununderstandable* response or no response at all and had it included in this category. This definition presents difficulties in the interpretation of the data.

In the arm there is no clear cut relationship to be seen between level of activity and learning score (Fig. 3), nor does the Kruskal-Wallis test approach significance. The non-monotonic relation of the activity level to learning scores exists in only certain localities.

Relation of change of level to learning performance.—As already stated, the level of activity in the head locations is lower during the information interval than during the response interval. This can be taken to mean that muscular activity falls off after a response is made. The question now is how this "decrease of tension" is related to the correctness of the

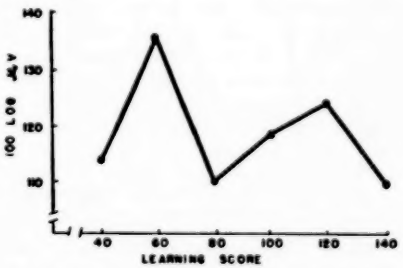

FIG. 3. The mean MAP values for the arm electrode as related to learning scores.

vocal response and to the learning performance. The amount of decrease was computed for each S and for right and wrong responses and the mean values (for the intervals on the learning performance scale) are plotted in Fig. 4. Neither curve departs significantly from a slope of zero by the Kruskal-Wallis test. A sign test (6, p. 57), however, shows that the decrease following a correct response is significantly different from zero ($P < .01$) but fails to show any significant departure for the decrease following an incorrect response. Since it would scarcely be proper on this evidence to affirm the null hypothesis in the one case and deny it in the other, the following comparison is made.

Differential behavior after right and wrong responses in relation to learning.— For each S a measure was prepared to represent the difference between the drop in MAP following a right response and the drop following a wrong response. The solid line in Fig. 5 is a plot of the group means of these measures. (The plot for right responses given in Fig. 4 is repeated here for reference.) The drop following right responses is not always greater; where the line falls below zero, it is less. To test the significance of what appears to be a systematic trend, the Kruskal-Wallis test was applied as before. Since, even by a chi-square interpretation $P < .02$, it is clear that there are differences among the learning scale groups, and that the differential behavior for right and wrong responses is related to learning scores. In view of this relation it is now possible to conclude that the behavior differs, as far as muscle tension is concerned, according to the correctness of the response. The difference however is not always of the same size, but varies with the learning score. Nor is it a monotonic function.

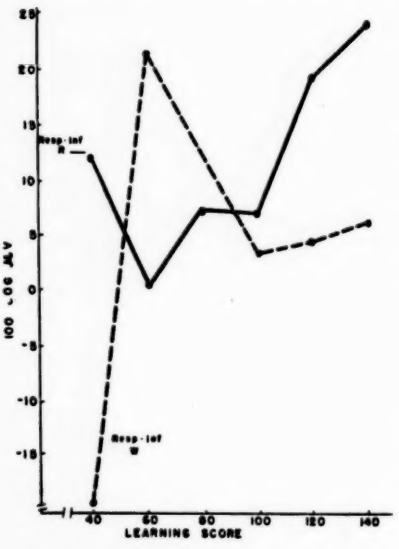

FIG. 4. The changes in MAP values (jaw and forehead combined) from response to information intervals for right and wrong responses as related to learning scores.

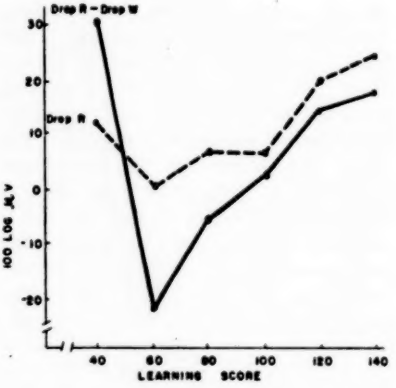

FIG. 5. The solid line represents the difference between each pair of points in Fig. 4. The dotted line is reproduced from the solid line of Fig. 4 for comparison.

Discussion

The experiment shows a muscular tension, somewhat localized, which is related to learning performance. The nature of the relation can be specified by describing the tension and tension changes in good, medium, and poor learners. The best performers have a high level of muscular activity in the head location, through all conditions. This tension nevertheless decreases substantially when *S* receives a confirmation of his verbal response. On the other hand there is little or no decrease in this activity when the signal in the information interval tells him that his response should have been something else. Good learners then are characterized by high activity level and differential responses to confirmation and correction. The moderately successful learners have the lowest activity level, little or no change when their anticipation is confirmed, and a notable decrease when their answers are corrected. They do have a differential response to confirmation and correction, but the difference is in quite the wrong direction, as judged by the performance of the good learners. The poor learners, like the best, have a high activity level; they also have, like the best learners, a difference between responses to confirmation and correction, with a greater decrease after confirmation. They differ from the best learners, however, in presenting a very large increase in muscular activity when they receive information that the last vocal response should have been something else. Of the various expressions considered, the behavior after a wrong response seems the most differentiated.

The non-monotonic relations which appear show that the tension-learning relation is more complex than usually thought. The relation of level of muscular activity to learning score is present, but opposite in the two parts of the curve. Decrease of activity after a correct response with little or no decrease after an error is favorable to learning, but very poor learners have a similar algebraic difference in the effects of confirmation and correction. The components differ, however. The upper end of the U-curves conforms with expectations from theories of learning and theories of tension effects.

Possibly in the upper segment of the curves the muscular activity recorded represents responses such as pronunciation of the syllable, vocally and subvocally. Such action would be expected to favor learning, especially as there is more of it following an incorrect response, which needs to be replaced, than following a correct response, which does not. In the lower range of the U-curve, on the other hand, muscular activities represented may be of another character, such as tendencies to escape, to combat *E*, to make subvocal ejaculations of surprise or dismay, or other indiscernable activity. The presence of these would account for a rise in the muscular activity when *S* is shown that he was wrong. It is quite possible that the more responses there are of this sort, the less learning is likely to occur.

If this interpretation is correct, the kind of muscular activity which is produced in a situation is of the greatest importance and it is an experimental challenge to identify the kind, if possible, from its internal characteristics.

In the foregoing interpretation the MAP condition in the information period has been taken as a change, usually a decrease from the level during the response interval. Considering only the data at hand it would also be possible to speak of the response period MAP as an increase over the information level. One would say, for example, that a good learner tends to show a greater increase when he is making a right response than he does when he is making an incorrect response. This interpretation, however, would include the statement that poor learners, in making a wrong response, show a decrease from their information period level. Because it would be an anomaly for *S*s to have a decreased MAP during a motor response, and because the response levels for right and wrong answers are closer together than are

the two corresponding information period levels, it seems better to use the response as the basis of comparison, and regard the information levels as changes from it, produced by the confirmation or correction which *S* receives. This is the interpretation given above.

Summary

Twenty-four male undergraduates were given 20 learning trials on a list of 9 nonsense syllables. The syllables were presented so that there was a 3-sec. intersyllable interval (response interval), and a 3-sec. exposure of each syllable (information interval). During learning, continuous action potential records were taken from the right forearm extensors, the jaw masseter, and the forehead frontalis. These records were analyzed for the first 10 right and the first 10 wrong responses of each *S* in terms of the maximum MAP's during the two 1.5-sec. intervals that made up the intersyllable (response) interval and the two 1.5-sec. intervals that made up the syllable exposure (information) interval. Also, each *S* was assigned a learning score, defined as the total number of right responses made during the 20 learning trials.

The data support the following conclusions:

1. The sum of jaw and forehead potentials, within *S*s, shows a significant non-linear relation with learning scores, the best and poorest learners having higher potentials than the mediocre ones.

2. Again summing jaw and forehead potentials, when a comparison was made between the change from response to information interval for the right responses as compared with this change for the wrong responses, the difference between the changes again shows a non-linear relation to learning scores, the differences being larger for the best and the poorest *S*s. Although it is not possible to determine whether the rights or the wrongs contribute more to this difference, there is a tendency for the wrongs to do so. In any case the differential behavior of MAP following right and wrong answers is related to learning.

3. No significant relations between arm potentials and learning scores could be discovered.

Several alternative interpretations of the data are suggested, with a preference for one which emphasizes the MAP changes following verbal responses.

References

1. Davis, R. C. Electromyographic factors in aircraft control. The relation of muscular tension to performance: Introduction. Rep. 55–122, Sch. Aviation Med, Randolph AFB, Dec., 1956.
2. Ellson, D. G., Davis, R. C., Burke, C. J., & Saltzman, I. J. *A report of research on detection of deception.* Contract N6onr-18011, Indiana Univer., 1952.
3. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–267.
4. Meyer, D. R. The interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50,** 204–220.
5. Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.
6. Walker, H. M., & Lev, J. *Statistical inference.* New York: Henry Holt, 1953.

(Received March 18, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

CUE AND CONTEXTUAL STIMULUS INTENSITY IN DISCRIMINATION LEARNING[1]

JOHN E. NYGAARD

University of Illinois[2]

The effect of the intensity of the conditioned or cue stimulus (CS) on the acquisition of a learned response has been the subject of numerous experimental studies. Although the results of these investigations have not been entirely consistent, the typical finding has been that an increase in intensity of the CS has been accompanied by an increase in response strength (**5**). Hull (**3, 4**) has accorded CS intensity a prominent role in his systematic formulation. In his theory both habit strength and reaction potential are increasing monotonic functions of CS intensity.

Recently, Logan (**6**) has shown that the effect of the intensity of the cue stimulus on response strength can be handled within Hull's theory without the necessity of introducing the stimulus intensity dynamism (V) postulate. Perkins (**8**) has made a very similar suggestion. Both of these writers point out that the correct response is reinforced only in the presence of the cue stimulus. Any occurrence of this response to stimuli in the contextual environment, to use Logan's term, is not reinforced, and inhibition to the contextual stimuli develops. In the situation in which the contextual stimuli are of a low or zero intensity, generalization of inhibition would be high to cue stimuli of low intensity and progressively less to stimuli of higher intensities. This would account for the results of experimental studies in which response strength is found to be an increasing function of the intensity of the cue stimulus. According to this formulation the stimulus intensity effect is not dependent upon the absolute intensity of the stimulus, as is the case with Hull's postulate, (V), but rather is dependent upon the relation between the intensity of the cue stimulus and the intensity of the contextual stimuli.

Perkins (**8**) reports as evidence for his interpretation a study in which he found response strength to increase as a function of the intensity of a light stimulus for a differentially conditioned group of rats, but found no such relation for a partial reinforcement control group. Bragiel and Perkins (**1**), in a study published while the experiment reported here was being conducted, found further support for the relative interpretation of stimulus intensity effects. In an instrumental jumping response, the mean latency for rats jumping to a white stimulus was significantly shorter than the mean latency for a group jumping to a black stimulus, when the background was black. However, when the background was white, latencies were significantly shorter to the black stimulus than to the white. The results of this study are difficult to evaluate, however, for the authors report that rats who jumped to stimuli of the same brightness as the background missed the stimulus door much more frequently than did the rats in the two groups in which the stimulus door and background contrasted. Thus, avoidance tendencies may have been confounded with stimulus differences.

[1] This paper constitutes a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate College of the University of Illinois, 1955. The writer wishes to express his appreciation to Dr. G. Robert Grice, who served as his advisor, for his valuable suggestions and criticisms.

[2] Now at Ohio University.

The present experiment was designed to determine the effect of differences in the cue and contextual stimulus intensities on the acquisition of an instrumental discrimination habit. Albino rats were trained on a black-white discrimination in which the contextual environment was either black, middle gray or white. Hull (**3**) has cited as evidence for his intensity dynamism postulate an unpublished black-white discrimination study in which rats trained to go to the white stimulus made fewer errors than those for which the positive stimulus was black. From Hull's interpretation one would predict the rats trained to a white positive stimulus would be superior in the acquisition of the black-white discrimination to rats trained to a black positive stimulus regardless of the brightness of the contextual environment. The position taken by Logan (**6**) and by Perkins (**8**), however, predicts that the white positive group will be superior only in the black environment, that there will be little or no difference in the middle gray environment, and that in the white environment the black positive group will be superior.

Method

Subjects.—The *S*s were 84 experimentally naive albino rats from the colony maintained by the Psychology Department of the University of Illinois. They ranged in age from 91 to 119 days at the beginning of the experiment. Twelve *S*s were discarded: five for failure to eat during preliminary training, three for failure to learn the door-opening habit, three for illness and one because of *E*'s error.

Apparatus.—The apparatus was a successive presentation discrimination apparatus almost identical to that used by Grice (**2**) for size discrimination. The main part of the apparatus consists of an alley 24 in. long, 5 in. wide, and 6 in. high. Vertical sliding doors were located in the middle and at each end of the alley. The alley was covered by lucite and mounted on a pivot so that either end could become the starting end or the reaction end. The second part of the apparatus held panels in which the stimulus doors were mounted. The panels and the stimulus doors were made of $\frac{1}{8}$-in. hardboard. The stimulus doors were 3×3 in., and were mounted flush with the panels. They were hinged at the top allowing *S* to push the door open and obtain a pellet of food placed in a cup attached to the back of the panel just below the door opening. The holder contained one stimulus panel at each end and was mounted on a pivot so either of the stimuli could be placed in position at the end of the alley. When the panel holding the stimulus door was in place against the end of the alley, the bottom of the stimulus door was $1\frac{1}{16}$ in. above the alley floor. Stops were placed so that if the stimulus were positive *S* could open the door and obtain a pellet of food but could not crawl behind the panel. If the stimulus were negative the door was locked, allowing *S* to push it $\frac{1}{16}$ in. but not allowing access to the food cup. The apparatus was illuminated by a 20-w. fluorescent lamp mounted 23 in. above the floor of the alley. This was the only illumination in the room.

Three such alleys were used in the experiment. One was painted a flat black, the second a middle gray, and the third a flat white. The middle gray was obtained by mixing the flat black and flat white paint. The panels holding the stimulus doors were painted the same color as the alley with which they were used. The stimulus doors were flat black or flat white. Reflectance measures of the black, gray and white were obtained from test patches using a General Electric Recording Spectrophotometer. The black had a reflectance of 3.5%, the gray 17.4%, and the white 87.2%. The ratio of black to gray was 1:4.97 and of gray to white 1:5.01. Brightness measures of the stimuli in the apparatus were made with a Macbeth Illuminometer. The absolute brightness level differed considerably in the three boxes, although the ratio of the black to the white stimulus in each box was the same.

Procedure.—The 72 *S*s were assigned at random to the six experimental groups with the limitation that six males and six females were assigned to each group.

All *S*s were placed on a 22.5-hr. food deprivation schedule seven days prior to preliminary door training. They were fed each day in individual feeding cages 8 gm. of Purina Lab Chow. This schedule was maintained throughout the experiment. Three days prior to the first training day each *S* was fed six pellets of Purina Lab Chow before the regular daily ration. These pellets weighed approximately .075 gm. Any *S* which was not eating the

pellets by the end of the third day was eliminated from the experiment.

Three days of door training preceded the discrimination training. This training was carried out in the alley to which *S* had been assigned for discrimination training. On the first day 10 pellets were placed in the food cup and the stimulus door was fastened open. The *S* was placed in the starting chamber and the center door was opened allowing *S* to enter the reaction end and eat the pellets. No *S* failed to eat the pellets within 45 min. The alley was then rotated, the other stimulus door moved into position and fastened open, and *S* was again allowed to enter the reaction chamber and eat 10 more pellets. Half of the *S*s in each group ate first from the black stimulus door and then from the white stimulus door. The other half ate first from white and then from black. During Days 2 and 3, *S*s were trained to push open the stimulus door and obtain a food pellet from the food cup; with 12 trials given each day. Black and white stimulus doors were presented on alternate trials. On Trial 1 of Day 2 the door was wide open. It was gradually closed on each trial. On Trials 9–12 of Day 3 it was completely shut. Three *S*s were eliminated for failure to eat within 4 min. on any of the final four trials.

Discrimination training was started on the following day. Two groups were given black-white discrimination training in each of the three alleys. One of the two groups in each alley was trained with the white stimulus positive and the other group with the black stimulus positive. The six groups are designated: B-B, black alley and black stimulus positive; B-W, black alley and white stimulus positive; G-B, gray alley and black stimulus positive; G-W, gray alley and white stimulus positive; W-B, white alley and black stimulus positive; W-W, white alley and white stimulus positive.

Discrimination training was carried out as follows: *S* was placed in the starting end; when *S* was oriented toward the center door, that door was opened allowing *S* to enter the reaction end. On trials to the positive stimulus the stimulus door was free to open so that *S* could obtain a pellet of food. On trials to the negative stimulus the door was locked, allowing it to be moved only $\frac{1}{8}$ in. and thus not providing access to the food cup. A pellet was always present in the food cup behind the negative stimulus door. Latency of response was recorded by a stop watch. The watch was started when the center door was opened and was stopped when *S* made a response to the stimulus door. A response was defined as a visible movement of the stimulus door made by *S*'s nose pushing against the door. On the first day of discrimination training a trial was terminated if no response was made within 4 min. On succeeding days trials were terminated at 45 sec. Following each trial the vertical door in front of the stimulus was lowered, the alley was pivoted, the appropriate stimulus put into position, and the next trial started. The *S*s were given 20 trials a day, 10 to the positive (P) and 10 to the negative (N) stimulus in the order: PNPPNNPNNPNPNNPPNPPN. Grice's (2) error score for successive discrimination training was used on each block of 10 trials. The *S*s were run until they reached the criterion of 18 out of 20 correct responses in two successive blocks of 10 trials, with 6 of the 10 responses to the negative stimulus having latencies greater than 10 sec.

Results

The acquisition curves for discrimination training are shown in Fig. 1. In plotting these curves the assumption was made that after reaching criterion an *S* continued to make 100% correct responses. As can be seen in the figure, marked differences between groups appeared on the first day of discrimination training.

Table 1 presents the median number of errors, exclusive of those made on criterion trials, for each of the six groups. The Mann-Whitney *U* test was used to test the difference between groups in each of the three alleys. It may be seen from the

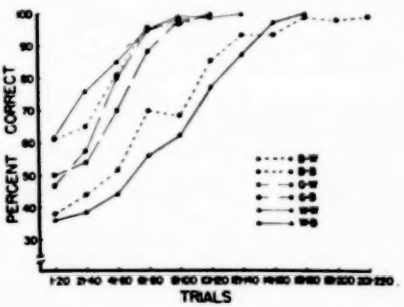

Fig. 1. Acquisition curves for each of the six groups.

TABLE 1

MEDIAN NUMBER OF ERRORS FOR EACH GROUP

Alley	Positive Stimulus		U
	Black	White	
Black	50.0	18.0	11.0*
Gray	24.0	28.0	53.5
White	14.0	62.0	0.0*

* $P < .05$.

table that the findings are consistent with the relational interpretation of stimulus intensity effects. Because the absolute intensities of the stimuli differed in the three alleys caution must be taken in comparing groups in different alleys. However, tests were run to determine whether the errors for groups in the gray alley differed significantly from those for the B-W and for the W-B groups. U tests show a significant difference at the .05 level between the W-B and G-B groups. The U for the B-W and G-W groups falls just short of the .05 level.

The number of trials to reach criterion, exclusive of criterion trials, and values of U are presented in Table 2. Differences between groups in each of the three boxes are consistent with results from the error scores. Differences between the gray box groups and the W-B and B-W group, however, are not significant.

TABLE 2

MEDIAN NUMBER OF TRIALS TO CRITERION FOR EACH GROUP

Alley	Positive Stimulus		U
	Black	White	
Black	110.0	55.0	26.5*
Gray	60.0	65.0	64.0
White	45.0	125.0	4.5*

* $P < .05$.

No analysis was made of the latency data; however, median latencies of each trial were plotted for all six groups.[3] An examination of these curves show that the latencies to the positive stimulus gradually decrease while those to the negative stimulus remain about the same or gradually decrease before increasing rather sharply. The major differences in the six groups are found in the latencies to the negative stimuli. In the W-B, B-W, G-B and G-W groups the median latencies to the negative stimuli begin to increase during Trials 40 to 60. However, these latencies in the B-B and W-W group continue to decrease beyond Trial 90, before increasing markedly.

DISCUSSION

The results give no evidence for an intensity dynamism effect in discrimination learning based on the absolute value of the intensity of the cue stimulus. The differences between groups are dependent on the relationship of the intensity of the cue stimulus to that of the contextual stimuli in the situation and, as such, agree with the predictions made from Logan's and Perkin's analysis of the problem.

The interpretation of generalization of extinction from contextual stimuli, however, does have some difficulty accounting for the fact that the W-B, B-W, G-W, and G-B groups reach criterion, i.e., extinguish to the negative stimulus, in approximately the same number of trials while the W-W and B-B groups take considerably longer. To do so it must be assumed that generalization of extinction is almost as great from gray to white or to black as it is from black to black or from

[3] These curves as well as the complete latency data may be found in the original dissertation (7).

white to white. This would require unusually flat generalization of extinction gradients over most of the range of brightnesses employed. Unfortunately, no generalization of extinction gradients for brightness stimuli are available.

Bragiel and Perkins' (1) study indicates that the orienting responses of rats are dependent on the stimulus-background relationship. It is possible that in the present study the cue-contextual intensity differences affect primarily the acquisition of the receptor orienting responses which are required for *S* to learn the discrimination. It is apparent from the acquisition curves that in the black and white alleys differences appear in the response latency to the black and white stimuli very early in training. If one postulates that the probability of the occurrence of the orienting response is greater to contrasting stimuli than to stimuli similar to their background, then the differences in acquisition of the discrimination may partially be accounted for. In the B-W, W-B, G-B and G-W groups during discrimination training a response to the contrasting stimulus is rewarded, whereas in the B-B and W-W groups a response to the contrasting stimulus is not followed by primary reward. Thus the orienting response comes to be associated with other stimuli in the situation more rapidly in those groups with a contrasting positive stimulus than those with a non-contrasting positive stimulus. Since the orienting response necessary to make the discrimination reaches a high level of probability of occurrences sooner in the B-W, W-B, G-W and G-B groups, the discrimination is learned more quickly. Initial differences in probability of occurrence of the orienting response can be accounted for, of course, in terms of the generalization of extinction interpretation. Whether the concept of the orienting response needs to be introduced at all, can be answered only by further experimental analysis.

Summary

This experiment investigated the effects of the brightness of the cue stimuli in relation to the brightness of the contextual stimuli on discrimination learning. The single stimulus presentation method was used. Six groups of 12 albino rats each were trained on a black-white discrimination problem. Two of the groups were trained on the discrimination in a black alley, two groups were trained in a gray alley and two groups were trained in a white alley. One of the groups in each box was trained with the white stimulus positive and the other group was trained with the black stimulus positive.

The results agree with a relative rather than an absolute interpretation of stimulus intensity effects. In the black alley the white positive group had significantly fewer errors and trials to reach criterion than did the black positive group. In the white alley the black positive group had significantly fewer errors and trials to reach criterion than did the white positive group. There was no significant difference between the groups in the gray alley.

References

1. Bragiel, R. M., & Perkins, C. C., Jr. Conditioned stimulus intensity and response speed. *J. exp. Psychol.*, 1954, **47**, 437–441.
2. Grice, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, **42**, 365–373.
3. Hull, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.
4. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.
5. Kessen, W. Response strength and conditioned stimulus intensity. *J. exp. Psychol.*, 1953, **45**, 82–86.
6. Logan, F. A. A note on stimulus intensity dynamism (*V*). *Psychol. Rev.*, 1954, **61**, 77–80.
7. Nygaard, J. E. Cue and contextual stimulus intensity in discrimination learning. Unpublished doctoral dissertation, Univer. of Illinois, 1955.
8. Perkins, C. C., Jr. The relation between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, **46**, 225–231.

(Received March 20, 1957)

Journal of Experimental Psychology
Vol. 55, No. 2, 1958

SHIFTS IN CONNOTATIVE MEANING OF WORDS AS A FUNCTION OF PREVIOUS RESTRICTIVE EXPERIENCE [1]

M. S. MAYZNER AND M. E. TRESSELT

Bell Telephone Laboratories and New York University

In the Method of Single Stimuli, *S* is presented with the members of a series of stimuli, one at a time, and asked to judge each with respect to a specified group of categories. Such a series of judgments leads to the production of an absolute scale or subjective ordering of the stimulus-range employed (**3**). When *S* is given practice on a limited portion of such a stimulus-range and then shifted to another or expanded portion of the range, a shift in the absolute scale first established is expected, and Helson (**1**), Johnson (**2**), Volkmann (**8**), and others (**3, 6, 7**) have elaborated the principles underlying such shifts.

Typically, in judgmental experiments involving shifts in absolute scales, the stimulus-range can be objectively ordered along some dimension such as mass, length, etc. If the stimuli used are words, however, certain difficulties arise in constructing an objectively ordered scale. For purposes of this study, a scale involving the frequency with which words are placed in a given conceptual category was developed. Such a word scale for any given conceptual category yields a list of words which vary in the frequency with which they would be judged as belonging to this category. That is, words at one end of this scale would be frequently judged as belonging to some given conceptual category, while words at the other end would be infrequently judged as belonging to this conceptual category.

The present experiment examines the properties of shifts of six absolute scales associated with the judgment of verbal stimulus material for six conceptual categories. Specifically, the following hypothesis is tested: The absolute scale of *S*, associated with practice given at one end of a frequency word scale (i.e., words showing high frequency of inclusion in a given concept) tends to become firmly anchored at this end and the shift or expansion needed to include items at the other end (i.e., words showing low frequency of inclusion in a given concept) is less likely to occur than when no such initial anchoring is present.

Method

Materials.—The stimulus material presented to *S* consisted of a list of words. These words were obtained and scaled from previous experimentation (**4, 5**) as follows: One hundred *Ss* were presented with a list of 600 words and were asked to check each word as belonging to none, 1, 2, 3, 4, 5, or 6 concepts. The six concepts were theoretical (T), economic (E), aesthetic (A), social (S), political (P), and religious (R).

By summing the number of checks for each word for the 100 *Ss* employed, 600 word profiles were obtained (**5**), which showed the frequency with which each word was judged as belonging to each of the six concepts. For example, the word profile for the word "science" showed frequency counts of 90, 21, 9, 23, 20, and 7 for the T, E, A, S, P, and R concepts, respectively, showing that 90 out of 100 *Ss* checked "science" as belonging to the theoretical concept, 21 out of

[1] The material in this paper was presented at the sixty-fourth annual convention of the American Psychological Association at Chicago, Illinois, in 1956.

TABLE 1

THE WORDS COMPOSING LISTS A AND B

List A*

Theoretical	Economic	Aesthetic	Social	Political	Religious
science	receipt	horizon	sympathy	vote	priest
reason	industry	poem	friend	senator	sermon
hypothesis	sale	easel	pet	nominate	reformation
method	expenditure	beauty	village	impeach	crucifix
analysis	capital	music	selfish	patronage	blessing
postulate	consolidated	blend	people	veto	spirit
experiment	margin	paint	affectionate	legislation	orthodox
synthesize	depreciation	verse	romance	representative	cathedral
prove	price	color	fraternity	campaign	temple
evidence	wages	picture	games	amendment	confession
inductive	distribution	museum	dear	constitution	faith
solution	rates	classical	gathering	congress	salvation
validity	money	artist	wedding	dictator	holy

List B**

1. table	14. heavy	27. penetrate	40. glaze	53. snap	66. glare
2. second	15. loud	28. shake	41. roll	54. sure	67. close
3. trunk	16. illumination	29. chamber	42. sink	55. twist	68. immerse
4. blanket	17. cast	30. tires	43. nail	56. dark	69. double
5. turn	18. coma	31. blade	44. rise	57. blunt	70. express
6. last	19. chair	32. spell	45. practice	58. confound	71. lock
7. substitute	20. rain	33. dry	46. organ	59. stilted	72. boil
8. blow	21. reveal	34. careful	47. pig	60. operation	73. contain
9. veil	22. moving	35. morsel	48. wind	61. jump	74. scalp
10. ass	23. lion	36. scalpel	49. discord	62. echo	75. remove
11. hand	24. mirror	37. death	50. dome	63. wooden	76. route
12. entry	25. dilatory	38. niche	51. back	64. prostrate	77. disclose
13. thunder	26. consider	39. rapt	52. listen	65. indent	78. ruin

* Words in List A were randomized for presentation to *S*.
** Words in List B are presented in the order in which they were presented to *S*.

100 *S*s checked it as belonging to the economic concept, etc.

These profiles were used to select two lists of words having properties as follows: List A (shown in Table 1) contains words having a frequency count which falls within the range from 70 to 100 for a given one of the six concepts and falls within the range from 0 to 30 for the remaining five concepts. For example, "science" showed a frequency count of 90 for the theoretical concept and frequency counts of 21, 9, 23, 20, and 7 for the remaining five concepts, all within the range from 0 to 30. List B (also shown in Table 1) contains words having frequency counts which fall within the range from 0 to 30 for all six concepts. For example, "table" showed frequency counts of 12, 29, 22, 26, 6, and 1 for the six concepts, T, E, A, S, P, and R, respectively. In brief, List A is comprised of words frequently judged as belonging to some one of the six concepts (with 13 words per concept), while List B is comprised of words infrequently judged as belonging to any of the six concepts, i.e., the two lists represent the two ends of a frequency-checking continuum or scale.

Procedure.—The *S*s of the present experiment were instructed to judge each word as to whether they felt it belonged to none, 1, 2, 3, 4, 5, or 6 of the concepts, as in the previous study. A control group of 156 *S*s was given the B list of words. An experimental group of 183 *S*s was given the AB list, i.e., the A list followed immediately by the B list, with the words numbered consecutively to give the appearance to *S* of a continuous list of words. The words of List B are presented in Table 1 in the order they were presented to *S*s. The words of List A are presented in Table 1 according to conceptual category, but were presented to *S*s in a random order. Both the B list, totaling 78 words, and the AB list, totaling 156 words, were presented to *S*s on mimeographed sheets along with instructions during regular class periods lasting

50 min. This provided ample time for the task. The *Ss* were instructed to check the words in sequence, to work carefully, but not to spend too much time on any given word. Both conditions, List B and List AB, were administered to each class, conditions being assigned to *Ss* at random. A total of 11 undergraduate psychology classes at New York University were employed and total testing was accomplished during a two-week period. None of the 339 *Ss* of the present experiment had been tested in the previous study from which the scaled stimulus material of this experiment was obtained.

Since the preceding design involves a comparison between checking behavior on the experimental List B and the control List B, any difference found could possibly be attributed to the mere presence of List A rather than to its frequency characteristics. Therefore a supplementary control group list is necessary whose frequency characteristics would be different from the present List A. Although such a supplementary control group (A'B) was run in connection with another study, in which checking behavior for words having different frequency characteristics was being related to Allport-Vernon value scores, its results would be relevant to the present study. Specifically, the frequency characteristics of words in List A' were in the range 0–30 for any three concepts and in the range 31–70 for the remaining three concepts. These 78 words of List A' were followed immediately by the 78 words of List B and were administered to 105 *Ss* attending undergraduate psychology courses at New York University, with the same instructions as in the present study, about one month earlier. Since the frequency characteristics of the words of List A' are more similar to the words of List B than the words of List A are to the words of List B, one would expect little or no decrement in List B checking in the A'B condition.

Results

Table 2 shows the concept spans, or mean number of words checked as belonging to each of the six conceptual categories, for the experimental and control groups. The results clearly indicate that for each of the six conceptual categories, the control group checked a significantly greater number of List B words than the experimental group. Thus, for the theoretical concept, the control group checked a mean number of 15.1 words as theoretical, while the experimental group checked a mean number of 9.4 words as theoretical. For the economic concept, the control group checked a mean number of 18.4 words as economic, while the experimental group checked a mean number of 11.3 words as economic. The remaining four concepts yield similar differences, as seen in Table 2, and the probability values associated with all six comparisons are highly significant, ranging from less than .02 for the social concept to less than .0001 for the theoretical and economic concepts.

Also of interest are the concept spans for the experimental group for

TABLE 2

Means and *SD*'s of the Concept Spans for the Experimental and Control Groups for the Six Conceptual Categories

Concept	List	Experimental Group (N = 183)		Control Groups (N = 156)*	
		Mean	*SD*	Mean	*SD*
Theor.	A	13.4	7.3		
	B	9.4	10.9	15.1 (14.1)	11.5 (16.4)
Econ.	A	17.5	8.1		
	B	11.3	11.2	18.4 (16.8)	13.2 (16.7)
Aesth.	A	15.8	7.5		
	B	9.5	9.6	13.5 (12.4)	10.4 (13.1)
Social	A	19.5	11.2		
	B	18.1	15.5	22.2 (22.9)	15.5 (18.9)
Polit.	A	18.0	8.1		
	B	9.0	10.1	12.4 (14.7)	11.3 (17.3)
Relig.	A	16.6	7.5		
	B	7.9	8.2	11.4 (11.6)	10.2 (14.7)

* Means and *SD*'s of List B for the supplementary control group (N = 105) are given in parentheses.

List A. Table 1 indicated that List A was composed of six groups of 13 words each, with the words of each group clearly belonging to a given conceptual category. Therefore, if the previous scaling technique is reliable, approximately 13 words should be checked for each conceptual category. The mean number of words checked for the six conceptual categories, namely T, E, A, S, P, and R, as shown in Table 2, are 13.4, 17.5, 15.8, 19.5, 18.0, and 16.6 words, respectively. These results on the average do not represent too great a departure from the predicted value of 13 words per conceptual category, and indicate that the words of List A were being judged and checked in accordance with the frequency scale values previously determined for those words.

It was noted in the procedure that a supplementary control group was necessary to insure that if checking decrements were found in the AB condition they could be attributed to the frequency characteristics of List A rather than to its mere presence. The results obtained for the supplementary control group for List B checking behavior are shown in parentheses in Table 2, and when

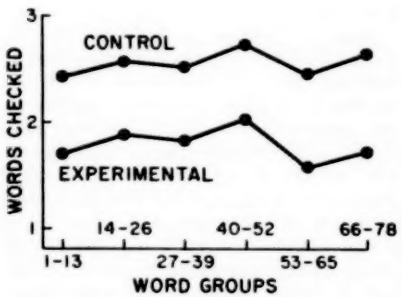

FIG 1. A plot of the average number of words checked for all six concepts combined on List B as a function of successive word groups for the control and experimental groups.

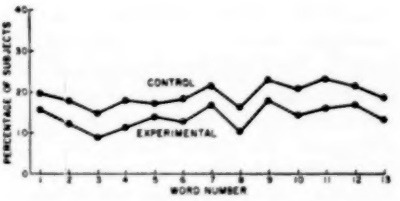

FIG. 2. A plot of the average percentage of *Ss* checking individual words 1–13 for all six concepts combined on List B for the control and experimental groups.

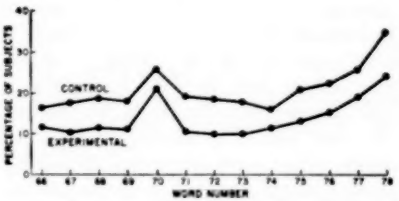

FIG. 3. A plot of the average percentage of *Ss* checking individual words 66–78 for all six concepts combined on List B for the control and experimental groups.

compared with the control group means of the present study show no significant differences, demonstrating that the frequency characteristics of the preceding words are crucial.

Having analyzed the general effect of List A on List B checking behavior, an analysis was made to determine how this checking decrement was distributed throughout the range of List B. The results are presented in three graphs. Figure 1 is a plot of the average number of words checked for all six concepts combined on List B, as a function of successive word-groups, for the control and experimental groups. More specifically, List B was broken down into six successive 13-word groups and the average number of words checked for all six concepts combined was calculated and plotted for each word-group, as shown. The most striking aspect of Fig. 1 is the general parallel-

ing of the curves, indicating that the checking decrements produced by List A are uniform throughout the entire range of List B, the decremental effect being just as strong on the last group of 13 words as on the first group of 13 words. The corresponding plots for each conceptual category when dealt with separately are of the same form as the composite shown in Fig. 1.

Figures 2 and 3 are plots of the average percentage of Ss who checked individual words, for all six concepts combined, on List B. Figure 2 shows the results for the first 13 words of List B and Fig. 3 shows the results for the last 13 words of List B. Again the most prominent feature is the general paralleling of the curves, with the checking decrement effect acting uniformly on individual words throughout the range of List B shown. To conserve space, the middle range of List B is not shown, nor the plots for each conceptual category when dealt with separately. The results correspond in form to those shown.

Discussion

The present investigation has attempted to deal with verbal conceptual or categorizing behavior using a judgmental model which typically is applied only to psychophysical problems. In making this application certain assumptions were made, as follows: (*a*) Categorizing a word as theoretical or not theoretical, economic or not economic, etc., is analogous to categorizing a weight as heavy or not heavy, or a line as long or not long, etc. (*b*) Words can be scaled along a dimension of frequency of inclusion in a given conceptual category as weights and lines can be scaled along dimensions of mass and length. In the light of the above assumptions, it was hypothesized that anchoring phenomena could be demonstrated with these verbal scales as with the more familiar psychophysical scales. In brief, the novel feature of this approach resides in treating verbal conceptualizing or categorizing behavior as a process in which a series of judgments are made on the stimuli being categorized.

The results of the present study clearly demonstrate anchoring effects of absolute scales associated with the categorizing of verbal stimuli. That is, conceptual categories, of the type dealt with here at least, do not have fixed boundaries but expand or contract in relation to contextual features and in accordance with the expansion and contraction of absolute scales as governed by anchoring effects. Specifically, the effect of List A, acting as an anchor, produces a significant decrement in List B checking behavior in the experimental group, in contrast to List B checking behavior in the control group, where the anchor is absent. The further finding that the anchoring effect produced by List A is uniform throughout the entire range of List B, as shown in Fig. 1, is somewhat surprising, for it might have been expected that as *S* was exposed to more and more List B words, the effect of List A would tend to weaken, and the experimental and control group curves would merge. That the curves do not merge leads to the conclusion that List A functions as a very strong anchor, whose effects extend uniformly and systematically through time.

Figures 2 and 3 provide one of the most significant findings of this study, namely, that the anchoring effect established by List A is so potent as to act uniformly on *each individual word* of List B, as shown by the parallel curves. The results shown in Fig. 1 for word groups, for example, could still have been obtained without the anchoring effect acting necessarily on each individual word. That this effect is found for all List B words, particularly when these words are especially difficult to categorize at all, would seem to indicate a remarkable stability and consistency in this type of verbal conceptualizing or cate-

gorizing behavior. It also raises questions concerning how persistent and stable this effect would be as a function of variations in length of both List A and List B. For example, "If the length of List B were increased would a point finally be reached where the experimental and control group curves would merge?" and, "What is the optimum length for List A to produce the greatest decrement in List B checking behavior?"

Summary

The present investigation attempted to analyze verbal conceptualizing or categorizing behavior in judgmental terms. Words previously scaled as to their frequency of inclusion in given conceptual categories served as the stimulus material and *Ss* were asked to judge each word as to whether it belonged to any of six conceptual categories. A control group of 156 *Ss* received a list of words (List B) and an experimental group of 183 *Ss* received List B preceded by an anchoring List A.

The results show that the anchoring effect of List A significantly reduced the number of stimulus items or words included under the conceptual categories in List B in the experimental condition in contrast to List B in the control condition. Further it was shown that this anchoring effect acts uniformly and systematically throughout the entire range of List B, producing uniform checking decrements on ndividual words.

References

1. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
2. Johnson, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.
3. McGarvey, H. R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol.*, 1943, No. 281.
4. Mayzner, M. S., & Tresselt, M. E. Concept span as a composite function of personal values, anxiety, and rigidity. *J. Pers.*, 1955, **24**, 20–33.
5. Pritchard, L., Mayzner, M. S., & Tresselt, M. E. The Mayzner-Tresselt word-concept check list. *Psychol. Newsltr.*, 1955, **7**, 27–33.
6. Rogers, S. The anchoring of absolute judgments. *Arch. Psychol.*, 1941, No. 261.
7. Tresselt, M. E. The effect of the experiences of contrasted groups upon the formation of a new scale of judgment. *J. soc. Psychol.*, 1948, **27**, 209–216.
8. Volkmann, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951.

(Received for early publication August 13, 1957)

NOW AVAILABLE

Publication Manual

of the American Psychological Association

1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. All scientists who are writing for publication will find the Publication Manual an indispensable guide.

Price, $1.00

Discounts for quantity orders over fifty copies

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.

ANNOUNCEMENT

JOURNAL OF EDUCATIONAL PSYCHOLOGY

This journal will now be published by the American Psychological Association. In 1958 it will become a bimonthly; issues will appear in February, April, June, August, October, and December. Contents include articles on problems of teaching, learning, and the measurement of psychological development.

All back issues and subscriptions up to and including the May 1957 issue are the property of Warwick and York, Inc., 10 East Centre Street, Baltimore 2, Maryland.

Subscription $8.00
(Foreign, $8.50)

Single Copies, $1.50

Direct new subscriptions and renewals to:

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.

KEEP POSTED

Read

CONTEMPORARY PSYCHOLOGY

A Journal of
Reviews
Criticism
Opinion

Can you keep posted on the ever-increasing number of books now published in your field—books on social and abnormal psychology, clinical psychology, consulting psychology, psychotherapy, personality, mental health?

Let *CP* help by bringing you each month selective reviews by competent reviewers of important books, controversial books, worthwhile books, useful books.

Read the Editor's page, CP SPEAKS, for news of the publishing world, comment on communication in general and in particular, discussion of reviewing and of criticism.

Subscription, $8.00
(Foreign, $8.50)

Single copy,
$1.00

Send subscription orders to:

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.